- ☐ These slides/notes represent only part of the course, and were accompanied by face-to-face explanations on white-board and additional topics / learning materials.
- ☐ In preparation of these slides I have also benefited from various books and online material.
- ☐ Some of the slides contain animations which may not be visible in pdf version.
- ☐ Corrections, comments, feedback may be sent to https://www.linkedin.com/in/naveedrazzaqbutt/

EE 302
Probabilistic Methods in Electrical Engineering

with

Dr. Naveed R. Butt

@

Jouf University

Introductions …

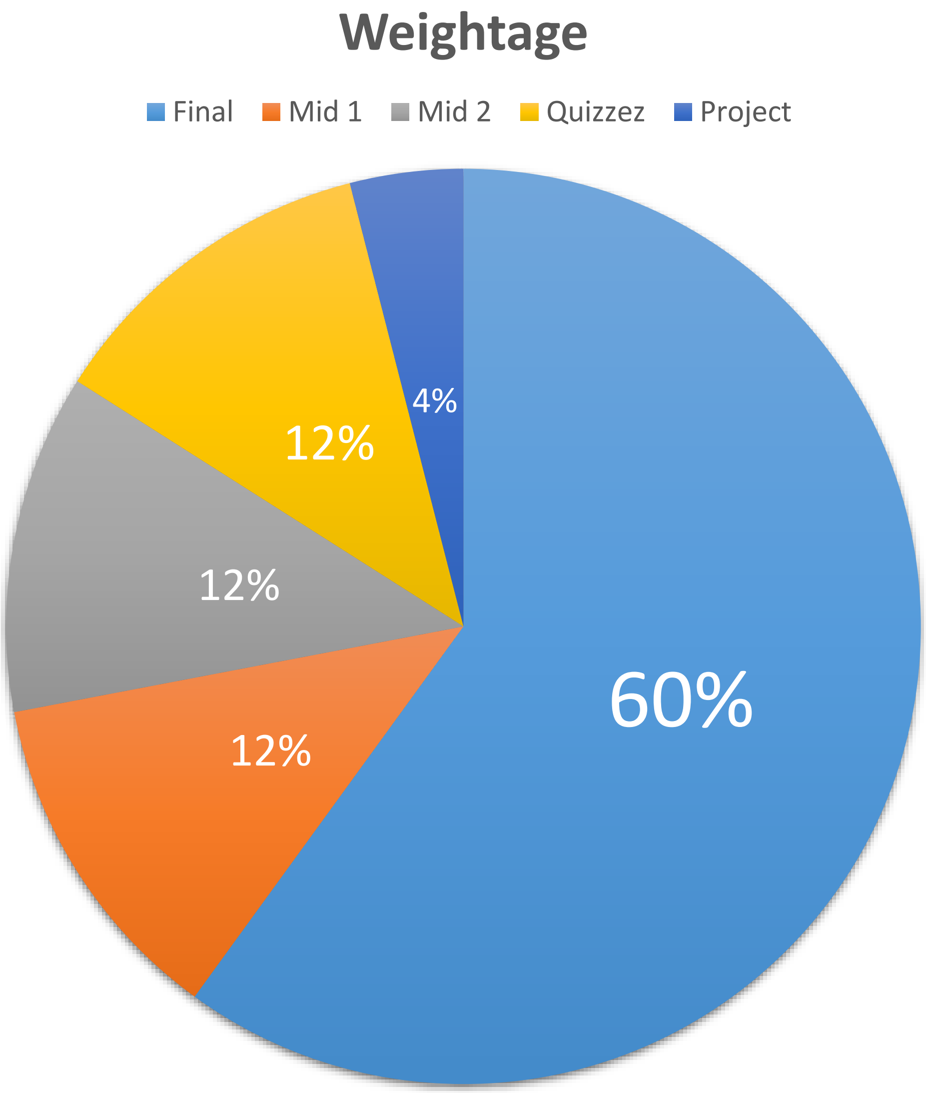
- Me
- You
- The Course

Important Business!!

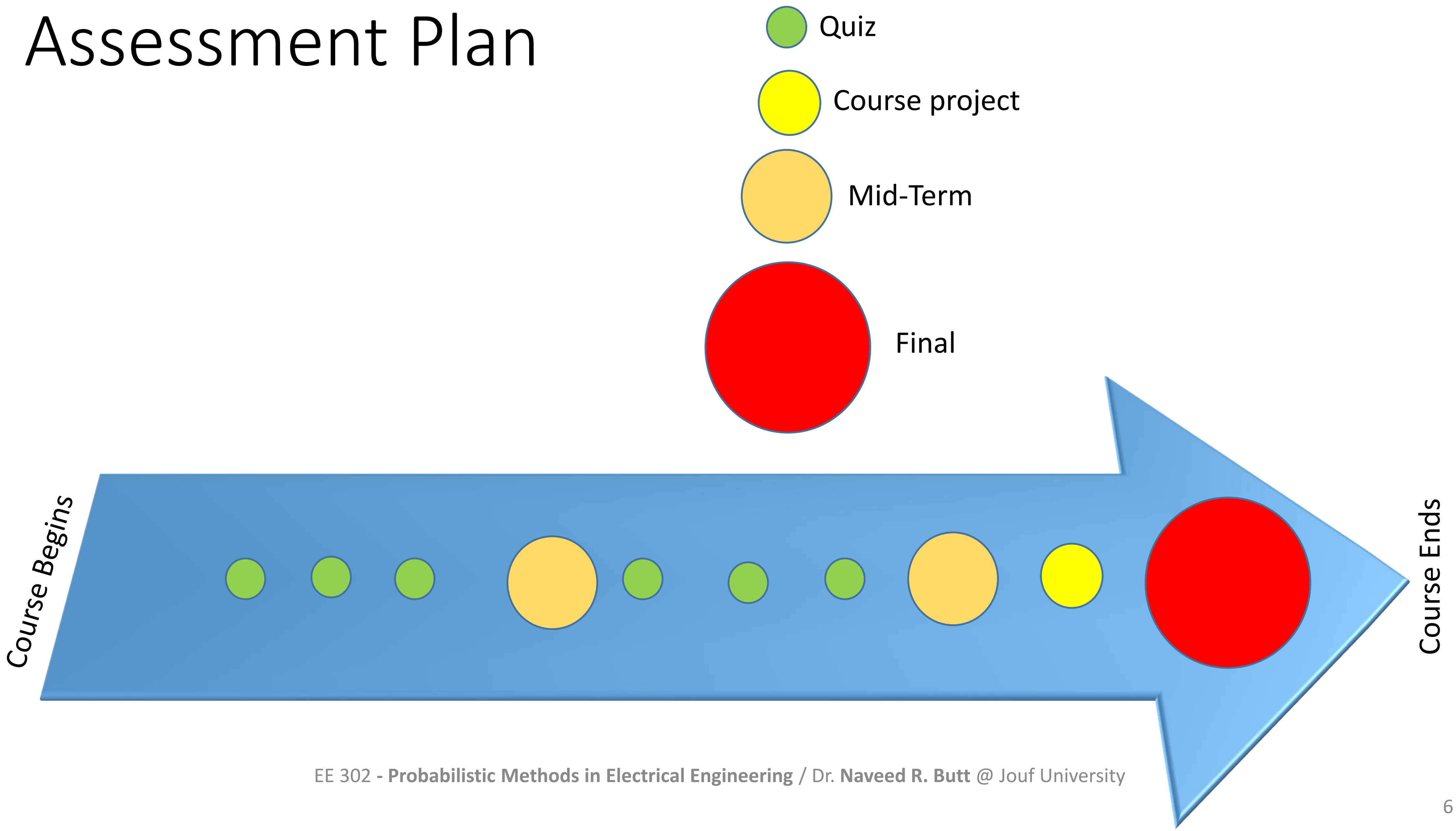
- 75% attendance is mandatory!
- Textbooks
  - Peyton Z. Peebles, JR, Probability, Random Variables and Random Signal Principles, 4th Edition, McGraw-Hill, 2002
  - D. C. Montgomery & G. C. Runge, Applied Statistics and Probability for Engineers, 6th Edition, Wiley, 2013
- Contact
  - nbutt@ju.edu.sa
  - office: 1140

Learning Plan

- **Lectures**
  - Help discover and grasp new concepts
- **Quizzes (six)**
  - Help prepare/revise each week's concepts
  - Keep you from lagging behind in course
- **Course Project**
  - Helps learn independent work & presentation
  - Prepares for final year project
- **Exams (Mid-1, Mid-2, Final)**
  - Help prepare entire course material



Assessment Plan



In this course we will see …

- What is “**probability**” and what are **its fundamental principles**
- **How can probabilities be assigned** to different kinds of events
- What are the **different types of random variables**
- What is meant by a **probability distribution** and what are the different **common probability distributions**
- How can we see the **link between various random events**
- What are **random processes** and some of **their characteristics**

Course Learning Objectives (CLOs)

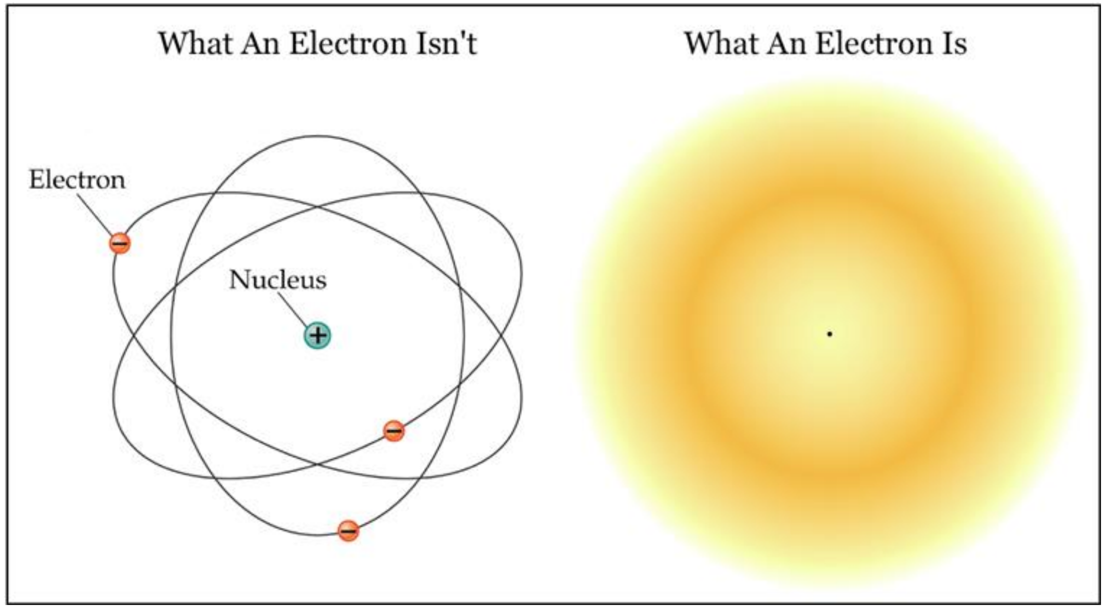
CLO #	Domain	Description	Assessment
CLO 1	Cognitive Skills	**Calculate** probabilities of events, joint probabilities, conditional probabilities using set operations and definition of probability. Justify valid and invalid probability assignments, and independence of events.	HW, Quiz, Mid, Final
CLO 2	Cognitive Skills	**Calculate** probability mass function parameters, moments and functions of discrete single and multiple random variables.	HW, Quiz, Mid, Final
CLO 3	Cognitive Skills	**Calculate** probability density function parameters, moments and functions of continuous single and multiple random variables.	HW, Quiz, Mid, Final
CLO 4	Cognitive Skills	**Analyze** random processes and effects of linear systems on random processes	HW, Final
CLO 5	Communication	**Demonstrate** the ability to research a topic related to probability and formally present the results	Project Presentation

What is “probability”?

What is “probability”?

- Probability is a “lack of knowledge”!
  - We know you are here today. We are sure.
  - But will you be here in the next lecture? We are not sure anymore! There is now a “lack of knowledge”
    - “perhaps”, “maybe”, “probably”

Why do we sometimes lack knowledge?



Why do we sometimes lack knowledge?

- Future
  - A dice you haven’t rolled yet
    - How can we know which number it will show!
- Too hard to collect all the information
  - Which places did you visit today?
    - It may be possible to have a drone camera follow you all the time. Then we will not have “lack of knowledge” about places you go to. But this is too hard a thing to do.
- Quantum randomness
  - Where’s the electron?
    - According to current consensus, processes and properties at quantum level are probabilistic by their very nature.

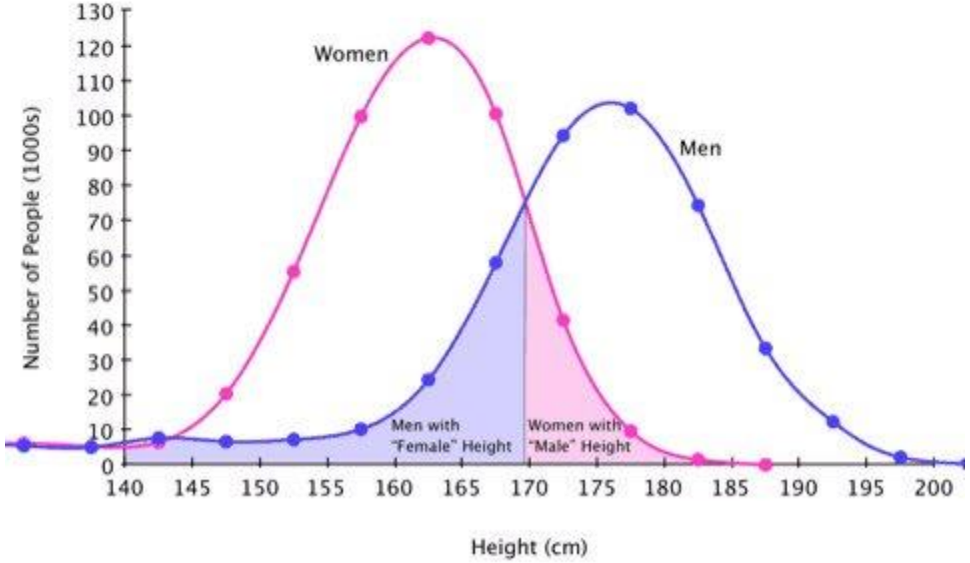
Why study probability?

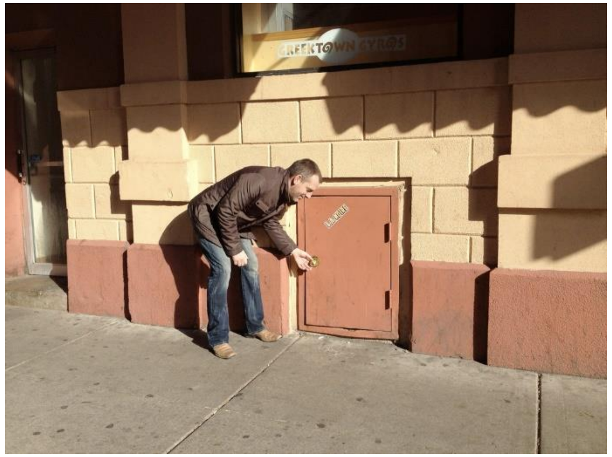
Why study probability?

- Height of the next student who enters the room.

Why study probability?

- Height of the next student who enters the room.





Why study probability?

- Height of the next student who enters the room.
  - There is lack of knowledge about it!
  - But that lack of knowledge is not "absolute"
    - We do *know something* about the heights of humans and can make some "guesses" based on whatever information we have (based on observation, experience, **statistics**)
  - Such guesses can help us design the height of the doorway (for instance).

Probability theory helps us make sense of an uncertain world!

- It helps us make ***smart guesses*** about uncertain events
- Based on the smart guesses we can ***plan, design, or take steps*** to better control the situation

Questions?? Thoughts??

EE 302
Probabilistic Methods in Electrical Engineering

with

Dr. Naveed R. Butt

@

Jouf University

A set of possibilities …

A set of possibilities …

- Suppose we **roll a dice**
- Before we can say something about the possibility of seeing a particular number, we have to determine what is possible and what is not.
- We need to define a ***set of possibilities*** for the dice rolling!

“Set” of possible outcomes

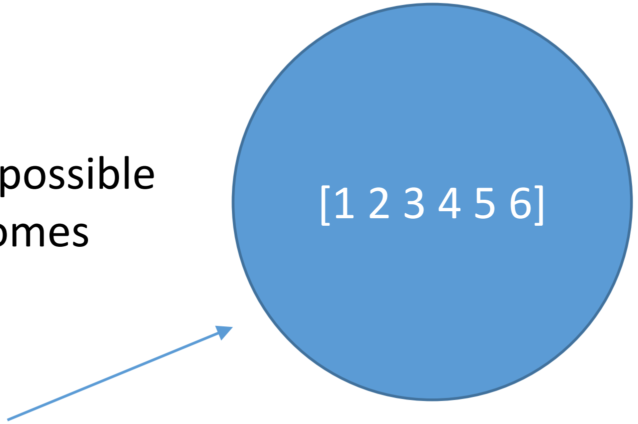
[1 2 3 4 5 6]

[0 7 8 …]

“Set” of impossible outcomes

A set of possibilities …

“Set” of possible outcomes



[1 2 3 4 5 6]

Universal Set or **Sample Space** (**S**) for the die rolling experiment

“Probability” and “Sets” go hand-in-hand!

- Understanding sets is an important part of understanding probability.

Definition: Set

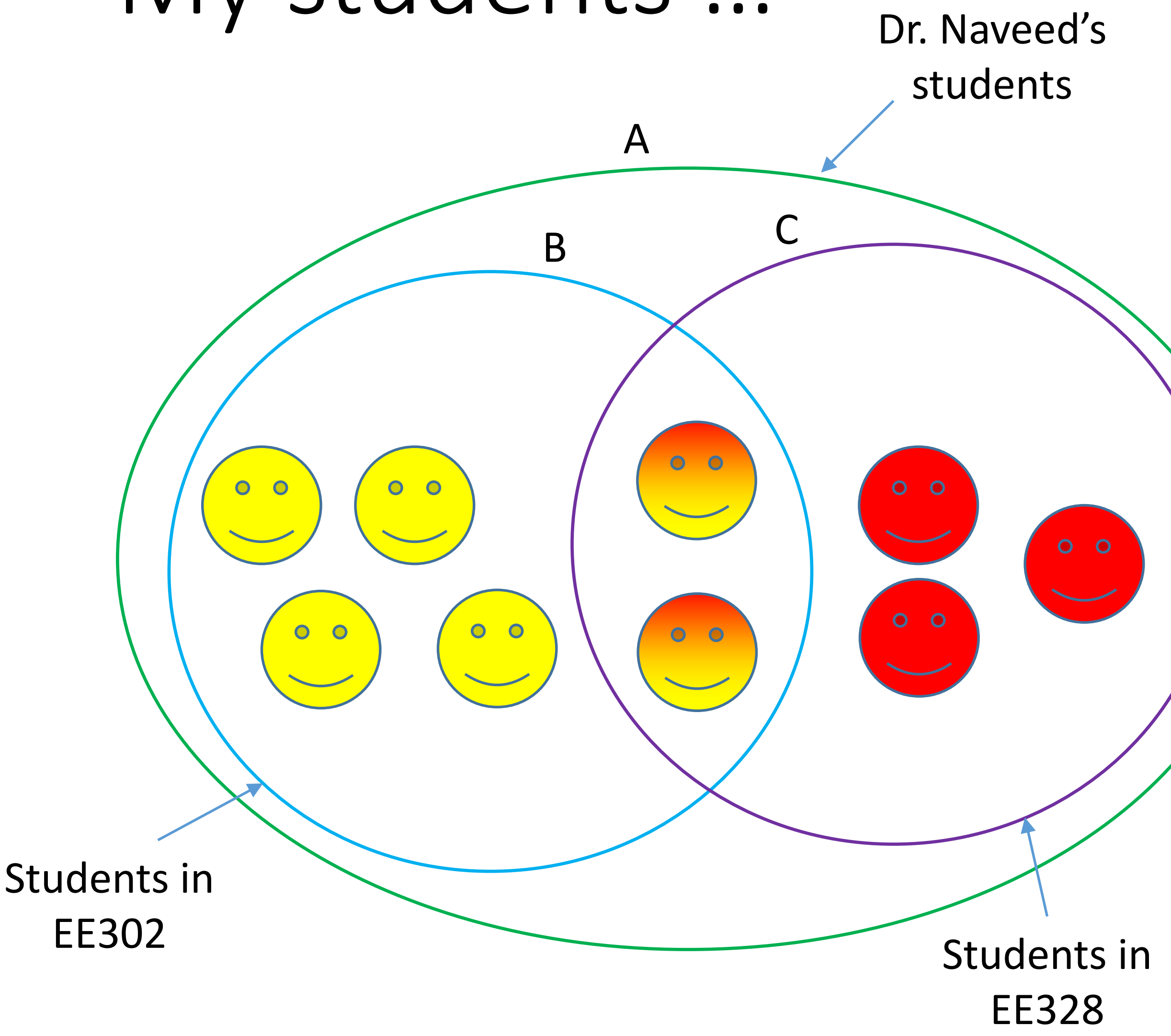
- A set is a collection of objects
- The objects are called “elements” of the set

$$A = \{a, b, c, d\}$$

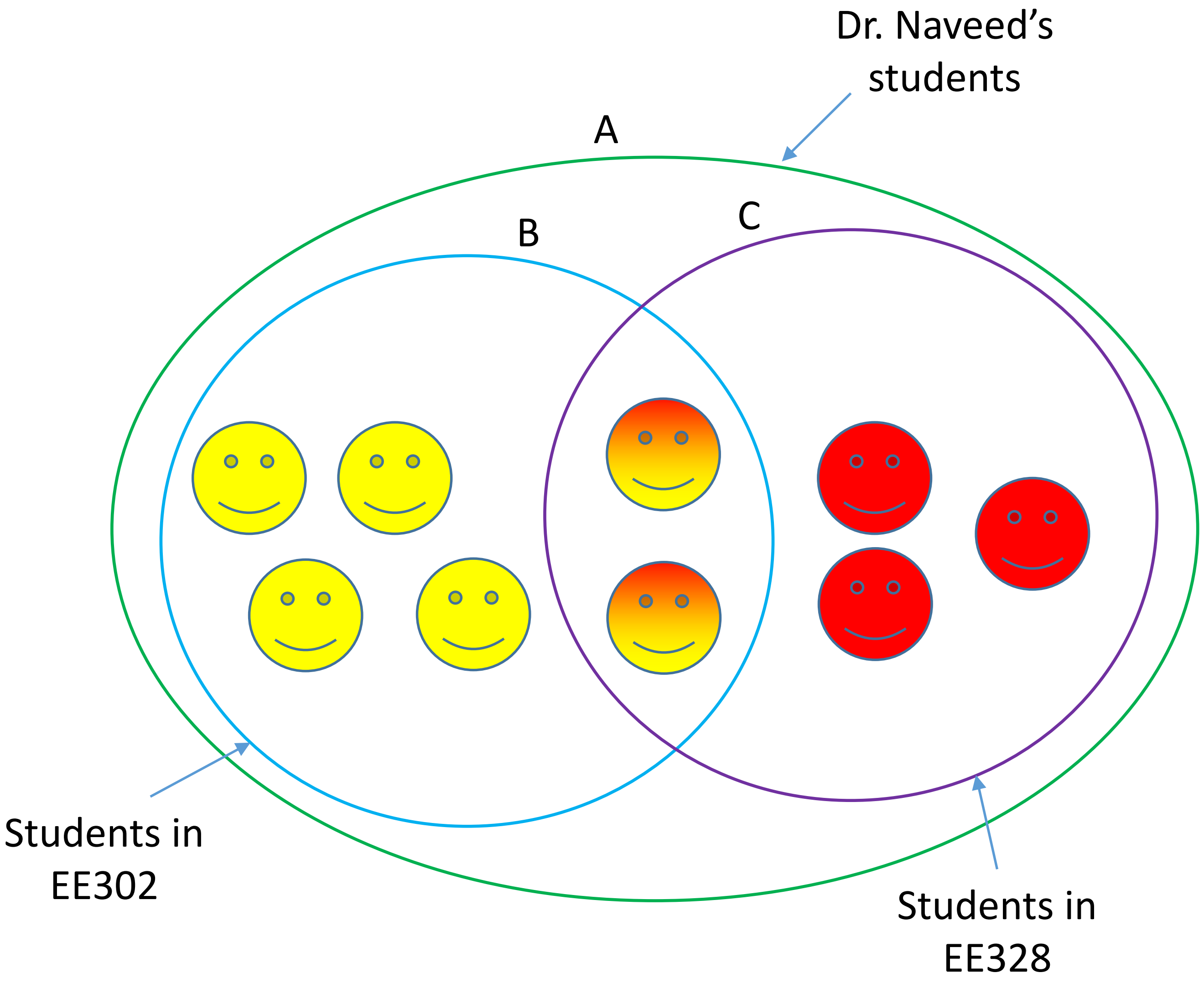
$$a \epsilon A$$

$$e \notin \mathrm{A}$$

My students …


Dr. Naveed's students
A
B
C
Students in EE302
Students in EE328

Basic set operations…



(assuming I am teaching only two courses this term)

B = { my students in EE302 }

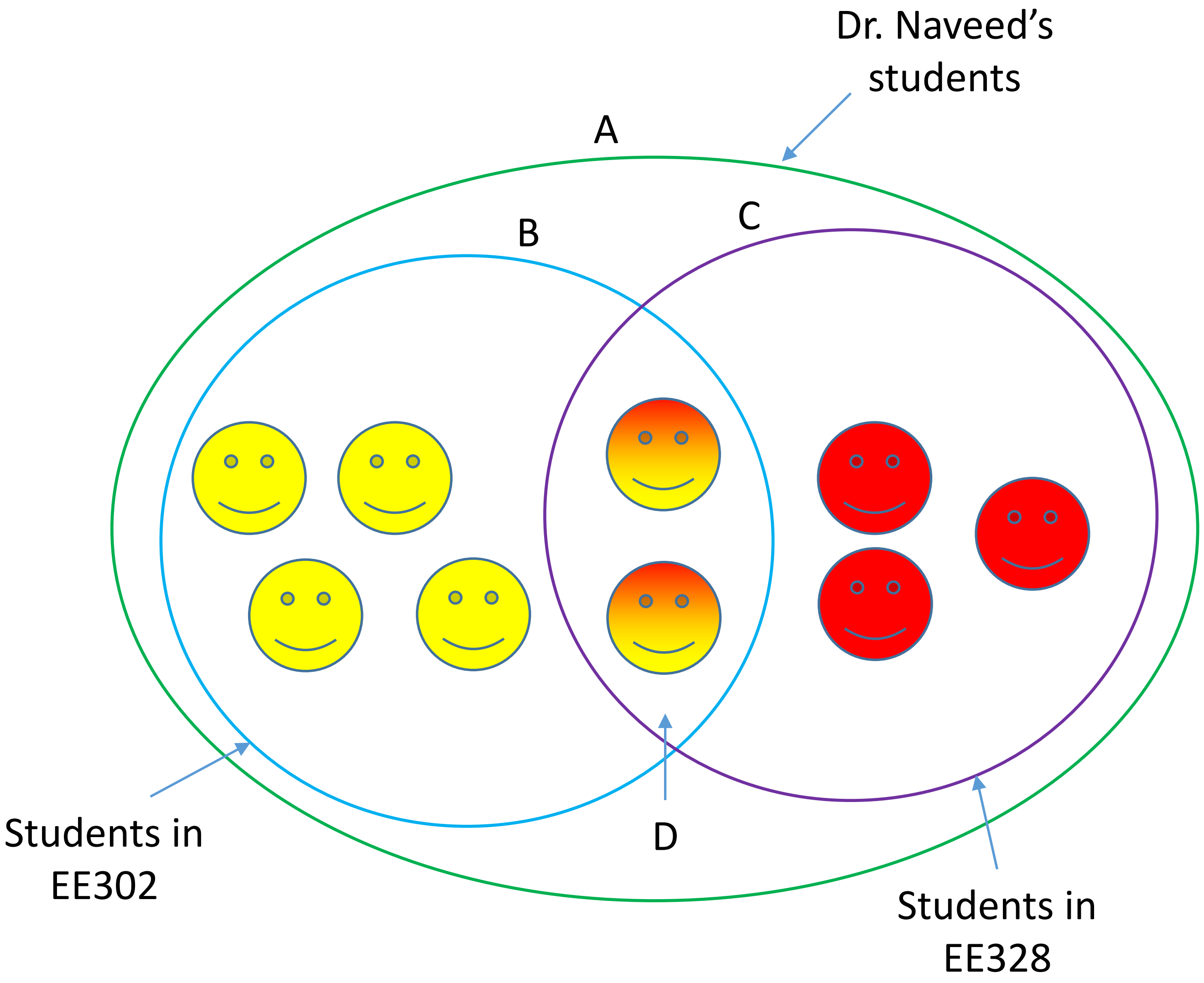
C = { my students in EE328 }

B U C = my students who are in EE302 ***or*** in EE328 = all of my students this term

B ∩ C = my students who are in both EE328 ***and*** EE302 (multicolor)

B – C = my students who are in EE302 but ***not in*** EE328 (yellow)

C – B = my students who are in EE328 but ***not in*** EE302 (red)



We also see that:

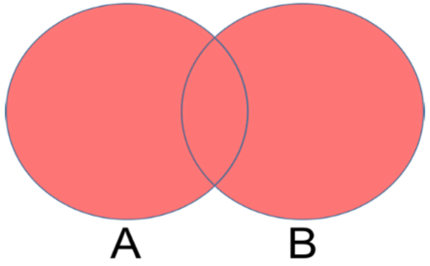
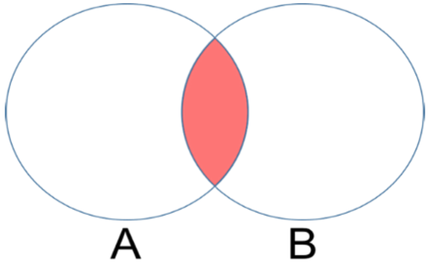
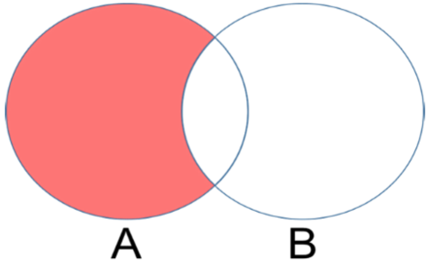
$\mathbf{D} \subseteq \mathbf{C}$ (all elements of D are also in C)

$\mathbf{B} \cup \mathbf{C} = \mathbf{C} \cup \mathbf{B}$

$\mathbf{B} \cap \mathbf{C} = \mathbf{C} \cap \mathbf{B}$

$\mathbf{B} - \mathbf{C} \neq \mathbf{C} - \mathbf{B}$

$\mathbf{A} - (\mathbf{B} \cup \mathbf{C}) = \emptyset$ (empty set)

Set Operation	Venn Diagram	Interpretation
Union		$A \cup B$, is the set of all values that are a member of A, or B, or both.
Intersection		$A \cap B$, is the set of all values that are members of both A and B.
Difference		$A \setminus B$, is the set of all values of A that are not members of B

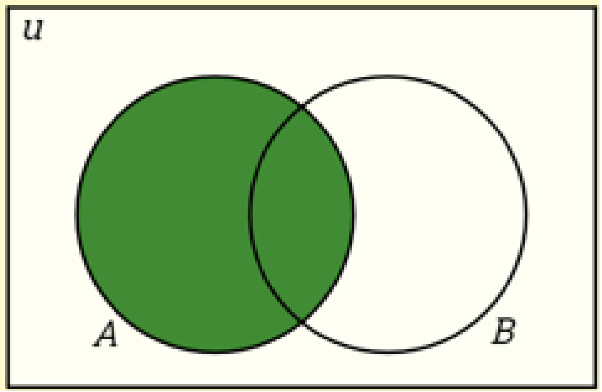
Some formal definitions …

- An “empty” set is a set with no elements
  - $\emptyset = \{\}$
- Two sets **A** and **B** are “**mutually exclusive**” or “**disjoint**” if they have no common elements
  - $\boldsymbol{A} \cap \boldsymbol{B} = \emptyset$
- A “universal set” is the largest or all-encompassing set of objects under discussion (usually denoted **S** or **U**)
- The complement of a set **A** is a set of all elements not in **A**
  - $\overline{\boldsymbol{A}}$ **= S - A**

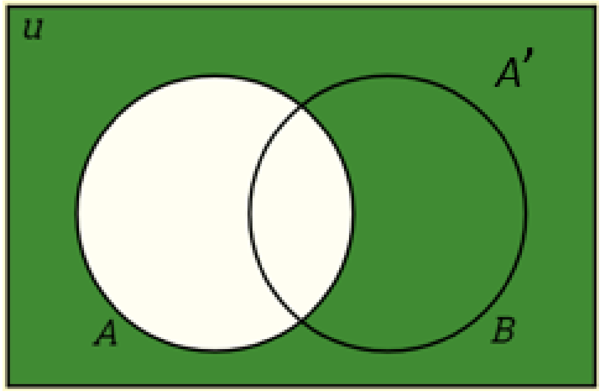
Some formal definitions …

- Two sets are “equal” if they are both subsets of each other (i.e., they have exactly the same elements)
  - $\boldsymbol{if}\ \boldsymbol{A} \subseteq \mathbf{B}\ \mathbf{and}\ \mathbf{B}\ \subseteq \mathbf{A}\ \mathbf{then}\ \mathbf{A} = \mathbf{B}$

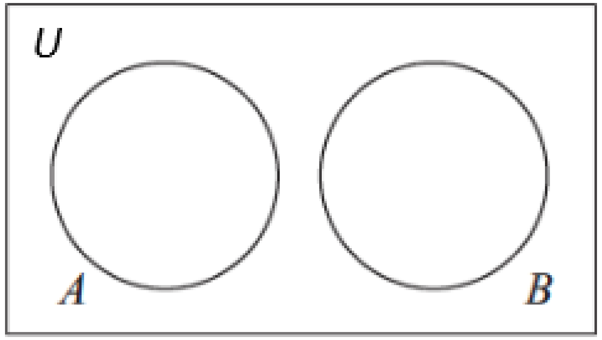
Set Operations and Venn Diagrams



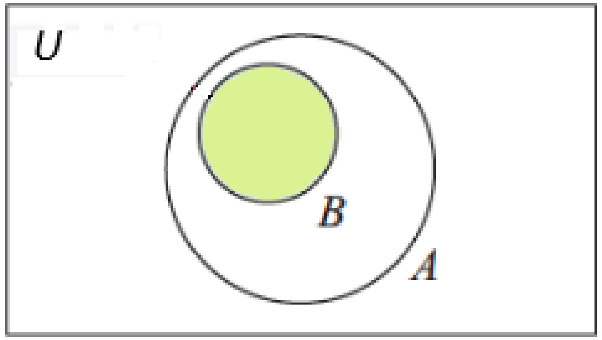
Set *A*



A′ the complement of *A*



A and *B* are disjoint sets



B is proper subset of *A* $B \subset A$

Three Ways of Defining a Set

- A set may be defined in **three** ways
  - **Tabular method**: we list (write) all the elements of the set
  - **Rule method**: we describe a rule for elements of the set
  - **Operation method**: we define the set as a result of some operation

Tabular	**A = {Abdallah, Mishari, Yazeed, Bandar, Basil, Badr, Abdullateef, Samir}**
Rule	**A = {students in EE302 second semester 2019}**
Operation	**C = B - A**

Notation

- Notation for Union and Intersection when more than two sets are involved
  - Recall that that when we add several variables we can denote the sum by using the symbol “sigma”

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n .$$

  - Similarly, union and intersection of several sets can be denoted as

$$C = A_1 \cup A_2 \cup \cdots \cup A_N = \bigcup_{n=1}^{N} A_n$$

$$D = A_1 \cap A_2 \cap \cdots \cap A_N = \bigcap_{n=1}^{N} A_n$$

Some properties of sets…

- Commutative, distributive and associative properties of union and intersection.

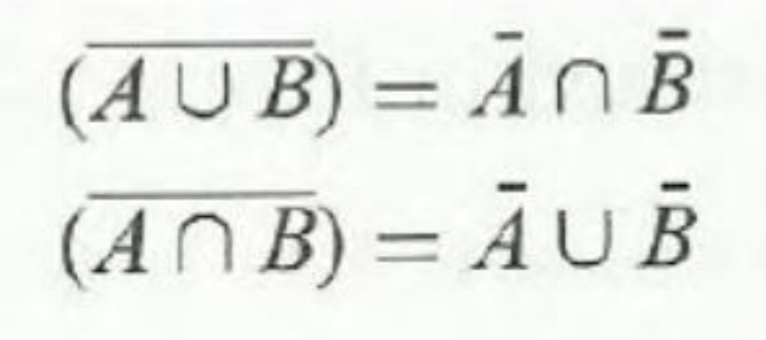
$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$$
$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$$

Some properties of sets…

- De Morgan’s Laws



$$(\overline{A \cup B}) = \bar{A} \cap \bar{B}$$
$$(\overline{A \cap B}) = \bar{A} \cup \bar{B}$$

Some properties of sets…

- Duality Principle
  - “if in an identity we replace unions with intersections, and intersections with unions, and also replace universal set (**S**) with empty set (∅), and empty set with universal set then identity remains valid”.

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

⇅

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

If one is valid, then so is the other!

Example Problems

Questions?? Thoughts??

EE 302
Probabilistic Methods in Electrical Engineering

with

Dr. Naveed R. Butt

@

Jouf University

In the last lecture, we saw that …

Quick Revision (1)

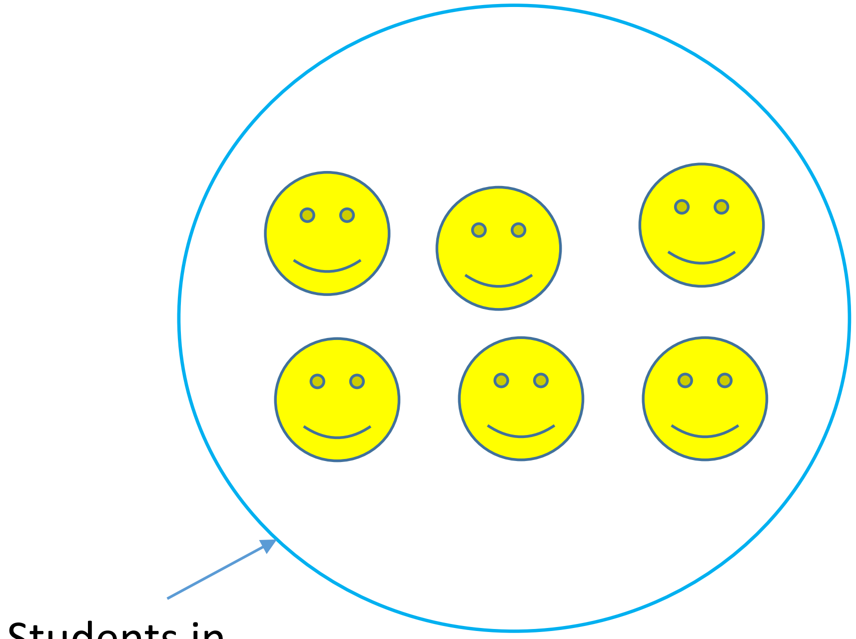
- A **set** is a **collection of objects**
- Objects in a set are called **elements** of the set
- We often **need sets to define probabilities**
- The set of all possible outcomes in a situation is called the **universal set** or **sample space** (denoted **S**)
- A set with no elements is called an **empty set** (denoted ∅ or **{}**)
- We saw how some sets are **related** to each other
  - **Subset**: when all the elements of set **A** can be found in set **B** $(\boldsymbol{A} \subseteq \mathbf{B})$
  - **Equality**: when two sets have exactly the same elements $(\boldsymbol{A} = \mathbf{B})$

$$A = \{a, b, c, d\}$$

$$a \in A$$

$$e \notin \mathrm{A}$$

Set = collection of objects



Students in
EE302

Quick Revision (1)

$A = \{a, b, c, d\}$

$a \in A$

$e \notin \mathrm{A}$

- A **set** is a **collection of objects**
- Objects in a set are called **elements** of the set
- We often **need sets to define probabilities**
- The set of all possible outcomes in a situation is called the **universal set** or **sample space** (denoted **S**)
- A set with no elements is called an **empty set** (denoted Ø or **{}**)
- We saw how some sets are **related** to each other
  - **Subset**: when all the elements of set **A** can be found in set **B** $(\boldsymbol{A} \subseteq \mathbf{B})$
  - **Equality**: when two sets have exactly the same elements $(\boldsymbol{A} = \mathbf{B})$

Quick Revision (1)

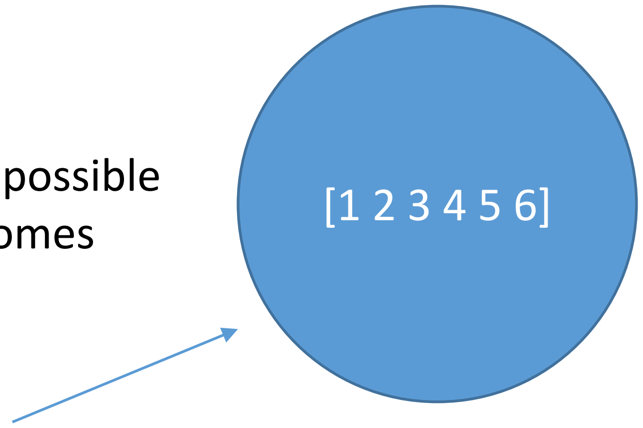
$A = \{a, b, c, d\}$

$a \in A$

$e \notin \mathrm{A}$

- A **set** is a **collection of objects**
- Objects in a set are called **elements** of the set
- We often **need sets to define probabilities**
- The set of all possible outcomes in a situation is called the **universal set** or **sample space** (denoted **S**)
- A set with no elements is called an **empty set** (denoted Ø or **{}**)
- We saw how some sets are **related** to each other
  - **Subset**: when all the elements of set **A** can be found in set **B** $(\boldsymbol{A} \subseteq \mathbf{B})$
  - **Equality**: when two sets have exactly the same elements $(\boldsymbol{A} = \mathbf{B})$

A set of possibilities …



“Set” of possible outcomes

[1 2 3 4 5 6]

Universal Set or Sample Space (**S**) for the die rolling experiment

Quick Revision (1)

$A = \{a, b, c, d\}$

$a \in A$

$e \notin \mathrm{A}$

- A **set** is a **collection of objects**
- Objects in a set are called **elements** of the set
- We often **need sets to define probabilities**
- The set of all possible outcomes in a situation is called the **universal set** or **sample space** (denoted **S**)
- A set with no elements is called an **empty set** (denoted ∅ or **{}**)
- We saw how some sets are **related** to each other
    - **Subset**: when all the elements of set **A** can be found in set **B** $(\boldsymbol{A} \subseteq \mathbf{B})$
    - **Equality**: when two sets have exactly the same elements $(\boldsymbol{A} = \mathbf{B})$

Quick Revision (1)

$A = \{a, b, c, d\}$

$a \in A$

$e \notin \mathrm{A}$

- A **set** is a **collection of objects**
- Objects in a set are called **elements** of the set
- We often **need sets to define probabilities**
- The set of all possible outcomes in a situation is called the **universal set** or **sample space** (denoted **S**)
- A set with no elements is called an **empty set** (denoted ∅ or **{}**)
- We saw how some sets are **related** to each other
  - **Subset**: when all the elements of set **A** can be found in set **B** $(\boldsymbol{A} \subseteq \mathbf{B})$
  - **Equality**: when two sets have exactly the same elements $(\boldsymbol{A} = \mathbf{B})$

Quick Revision (1)

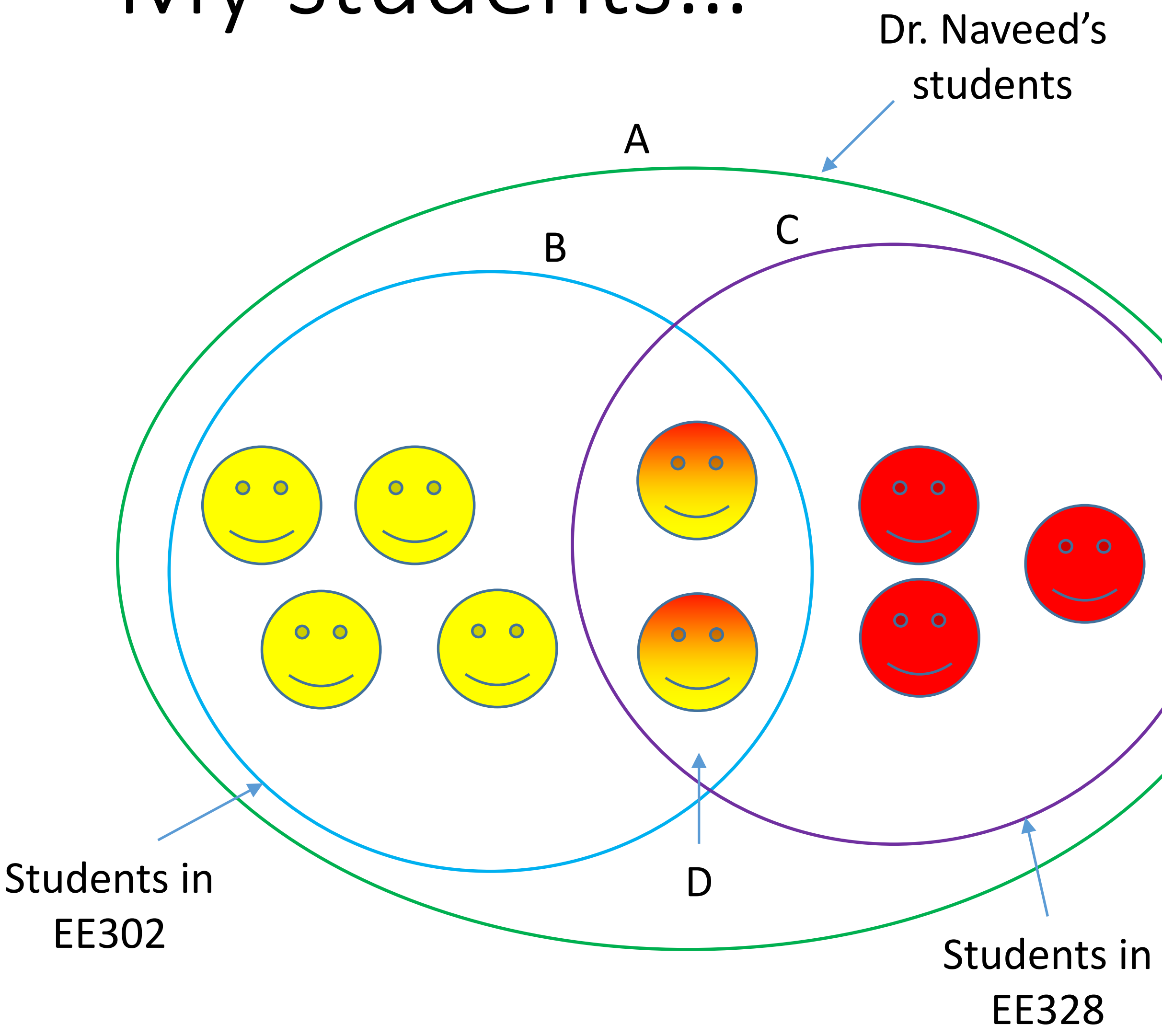
$A = \{a, b, c, d\}$

$a \in A$

$e \notin \mathrm{A}$

- A **set** is a **collection of objects**
- Objects in a set are called **elements** of the set
- We often **need sets to define probabilities**
- The set of all possible outcomes in a situation is called the **universal set** or **sample space** (denoted **S**)
- A set with no elements is called an **empty set** (denoted Ø or **{}**)
- We saw how some sets are **related** to each other
  - **Subset**: when all the elements of set **A** can be found in set **B** $(\boldsymbol{A} \subseteq \mathbf{B})$
  - **Equality**: when two sets have exactly the same elements $(\boldsymbol{A} = \mathbf{B})$

My students…



$\mathbf{B} \subseteq \mathbf{A}$ (all elements of B are also in A)

Quick Revision (2)

- We saw how we can **check equality** of two sets
  - $\boldsymbol{if\ A \subseteq \mathrm{B}\ \mathrm{and}\ \mathrm{B}\ \subseteq \mathrm{A}\ \mathrm{then}\ \mathrm{A} = \mathrm{B}}$
- We learned that two sets are completely different (**mutually exclusive** or **disjoint**) if they have no elements in common
  - Check for mutual exclusiveness : $\boldsymbol{A}\ \cap \boldsymbol{B} =\ \emptyset$
- We also saw some **operations** we can perform on sets
  - **Union**: $\boldsymbol{A} \cup \boldsymbol{B}$ take all the elements of the sets **A** and **B**
  - **Intersection**: $\boldsymbol{A} \cap \boldsymbol{B}$ = take only those elements which are common between **A** and **B**
  - **Difference**: $\boldsymbol{B} - \boldsymbol{A}$ = remove from **B** all those elements which are also present in **A**
  - **Complement**: $\overline{\boldsymbol{A}}$ **=** $\boldsymbol{S} - \boldsymbol{A}$ (remove all elements of set **A** from universal set **S**)
- Finally, we discussed some useful **properties of set operations**
  - **Commutative**, **Distributive**, **Associative**
  - **De Morgan's Laws**
  - **Duality Principle**

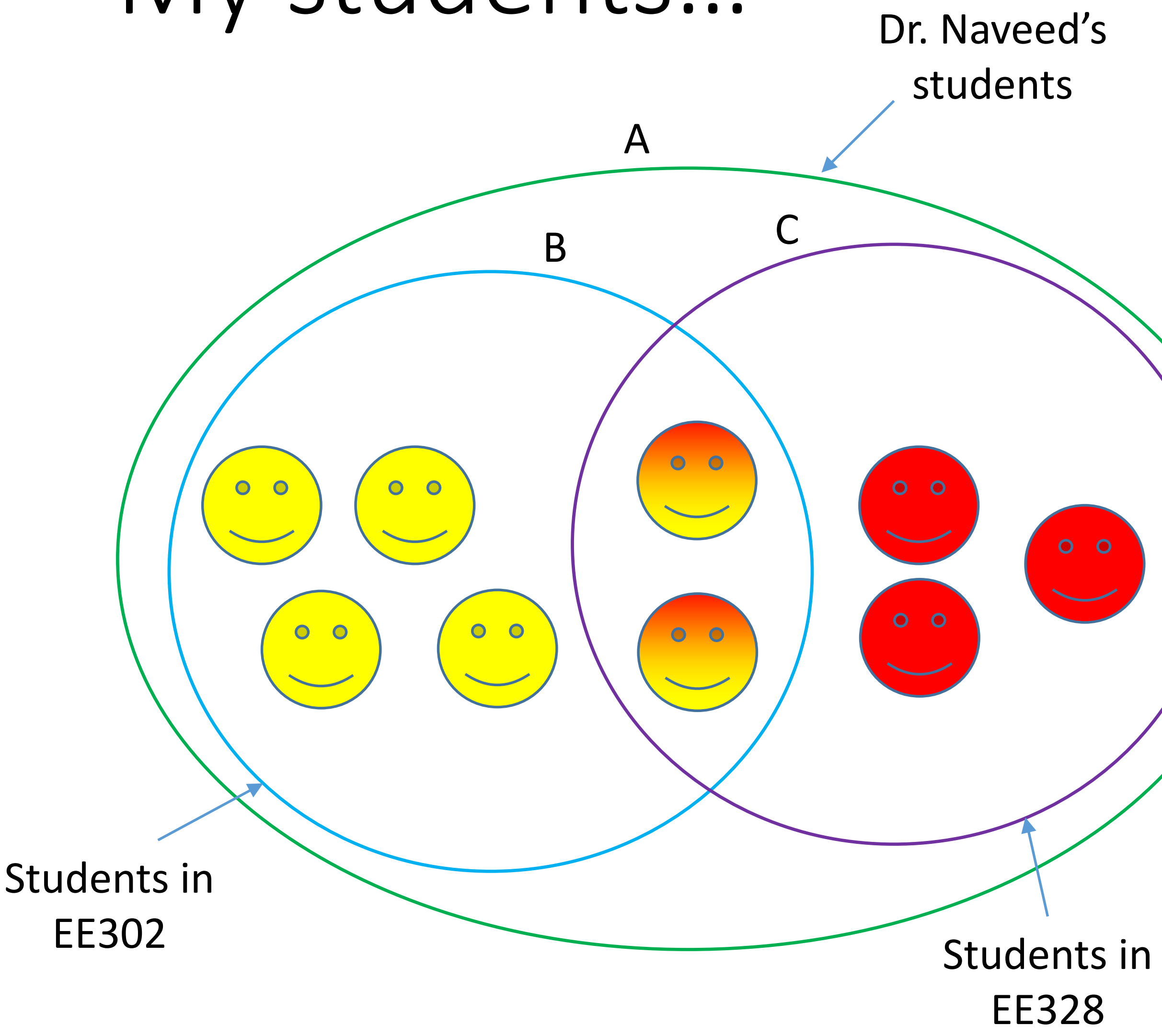
Quick Revision (2)

- We saw how we can **check equality** of two sets
  - $if\ A \subseteq \mathbf{B}\ \mathbf{and}\ \mathbf{B}\ \subseteq \mathbf{A}\ \mathbf{then}\ \mathbf{A} = \mathbf{B}$
- We learned that two sets are completely different (**mutually exclusive** or **disjoint**) if they have no elements in common
  - Check for mutual exclusiveness : $\boldsymbol{A} \cap \boldsymbol{B} = \emptyset$
- We also saw some **operations** we can perform on sets
  - **Union**: $\boldsymbol{A} \cup \boldsymbol{B}$ take all the elements of the sets **A** and **B**
  - **Intersection**: $\boldsymbol{A} \cap \boldsymbol{B}$ = take only those elements which are common between **A** and **B**
  - **Difference**: $\boldsymbol{B} - \boldsymbol{A}$ = remove from **B** all those elements which are also present in **A**
  - **Complement**: $\overline{\boldsymbol{A}}$ **=** $\boldsymbol{S} - \boldsymbol{A}$ (remove all elements of set **A** from universal set **S**)
- Finally, we discussed some useful **properties of set operations**
  - **Commutative**, **Distributive**, **Associative**
  - **De Morgan's Laws**
  - **Duality Principle**

Quick Revision (2)

- We saw how we can **check equality** of two sets
  - $if\ A \subseteq \mathbf{B}$ **and** $\mathbf{B} \subseteq \mathbf{A}$ **then** $\mathbf{A} = \mathbf{B}$
- We learned that two sets are completely different (**mutually exclusive** or **disjoint**) if they have no elements in common
  - Check for mutual exclusiveness : $\boldsymbol{A} \cap \boldsymbol{B} = \emptyset$
- We also saw some **operations** we can perform on sets
  - **Union**: $\boldsymbol{A} \cup \boldsymbol{B}$ take all the elements of the sets **A** and **B**
  - **Intersection**: $\boldsymbol{A} \cap \boldsymbol{B}$ = take only those elements which are common between **A** and **B**
  - **Difference**: $\boldsymbol{B} - \boldsymbol{A}$ = remove from **B** all those elements which are also present in **A**
  - **Complement**: $\overline{\boldsymbol{A}}$ **=** $\boldsymbol{S} - \boldsymbol{A}$ (remove all elements of set **A** from universal set **S**)
- Finally, we discussed some useful **properties of set operations**
  - **Commutative**, **Distributive**, **Associative**
  - **De Morgan's Laws**
  - **Duality Principle**

My students…



B = { my students in EE302 }
C = { my students in EE328 }
B U C = my students who are in EE302 ***or*** in EE328 = all of my students this term
B ∩ C = my students who are in both EE328 ***and*** EE302 (multicolor)
B – C = my students who are in EE302 but ***not in*** EE328 (yellow)
C – B = my students who are in EE328 but ***not in*** EE302 (red)

Quick Revision (2)

- We saw how we can **check equality** of two sets
  - $\boldsymbol{if\ A \subseteq \mathrm{B\ and\ B} \subseteq \mathrm{A\ then\ A = B}}$
- We learned that two sets are completely different (**mutually exclusive** or **disjoint**) if they have no elements in common
  - Check for mutual exclusiveness : $\boldsymbol{A} \cap \boldsymbol{B} = \emptyset$
- We also saw some **operations** we can perform on sets
  - **Union**: $\boldsymbol{A} \cup \boldsymbol{B}$ take all the elements of the sets **A** and **B**
  - **Intersection**: $\boldsymbol{A} \cap \boldsymbol{B}$ = take only those elements which are common between **A** and **B**
  - **Difference**: $\boldsymbol{B} - \boldsymbol{A}$ = remove from **B** all those elements which are also present in **A**
  - **Complement**: $\overline{\boldsymbol{A}}$ **=** $\boldsymbol{S} - \boldsymbol{A}$ (remove all elements of set **A** from universal set **S**)
- Finally, we discussed some useful **properties of set operations**
  - **Commutative, Distributive, Associative**
  - **De Morgan's Laws**
  - **Duality Principle**

Commutative Property

$$A \cap B = B \cap A$$
$$A \cup B = B \cup A$$

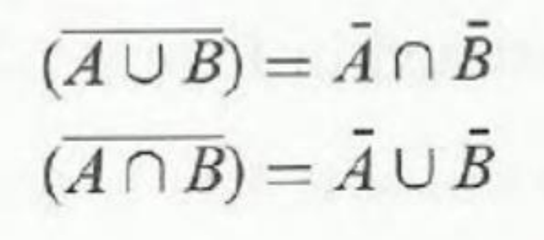
Distributive Property

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Associative Property

$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$$
$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$$

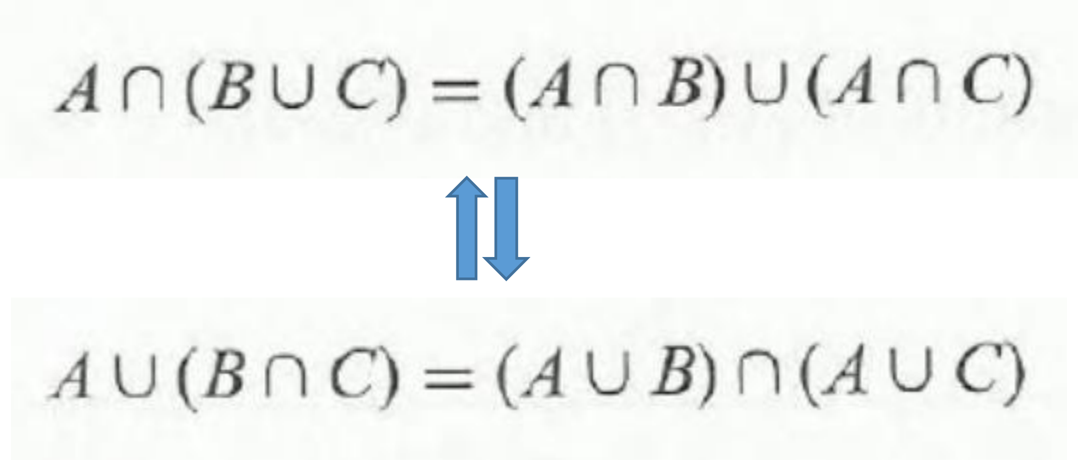
De Morgan's Laws



$$(\overline{A \cup B}) = \bar{A} \cap \bar{B}$$

$$(\overline{A \cap B}) = \bar{A} \cup \bar{B}$$

Duality Principle



$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$\Updownarrow$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

*"if in an identity we replace unions with intersections, and intersections with unions, and also replace universal set (**S**) with empty set (∅), and empty set with universal set then identity remains valid".*

Today …

- We will go through some examples
- And, see how we can assign probabilities to random events
  - What is the process?
  - What are the conditions?

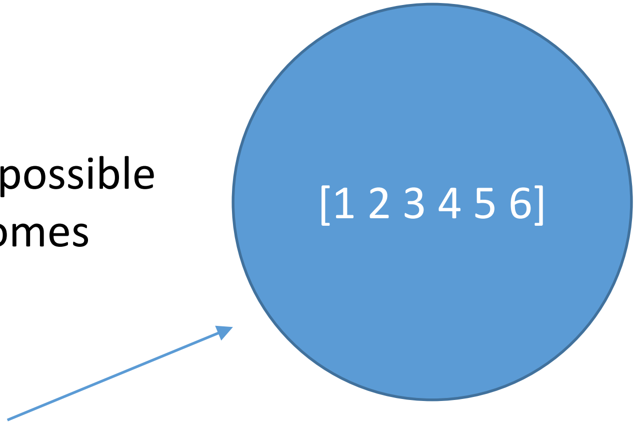
Definition: Random Event

- An outcome (or result) we are not sure about.
  - Example: when we toss a coin the occurrence of a head or tail is a random event.

Defining “Probability” of a Random Event

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total \# of outcomes}}$$

A set of possibilities …



“Set” of possible
outcomes

Universal Set or **Sample Space**
(**S**) for the die rolling experiment

How do we assign probabilities to random events?

- What is the process?
- What are the conditions?

Let’s play …

- Assume there is a bag with only one ball in it which is yellow
  - You put your hand in the bag, and without looking, take a ball out and note its color
  - What are the possible outcomes?

Set of all the possible outcomes → $S = \{\text{Yellow}\}$

Let’s play …

- What is the probability that the ball you draw is yellow?

Outcome of our interest

A = {Ball drawn is Yellow}

Event

Set of all the possible outcomes

S = {Yellow}

Universal Set

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total \# of outcomes}}$$

P(A) = P(yellow ball) = 1

100% (sure)

Let’s play …

- Now assume there are two balls in the bag, yellow and red
  - You put your hand in the bag, and without looking, take a ball out and note its color
  - What are the possible outcomes?

Set of all the possible outcomes → $S = \{\text{Yellow, Red}\}$

Let’s play …

- What is the probability that the ball you draw is yellow?

Outcome of our interest

A = {Ball drawn is Yellow}

Event

Set of all the possible outcomes

S = {Yellow, Red}

Universal Set

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total \# of outcomes}}$$

P(A) = P(yellow ball) = 1/2

50% chance

Let's play …

- What is the probability that the ball you draw is yellow?

Outcome of our interest

A = {Ball drawn is white}

Event

Set of all the possible outcomes

S = {Yellow, Red}

Universal Set

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total \# of outcomes}}$$

P(A) = P(white ball) = 0

impossible

Let’s play …

- What is the probability that the ball you draw is yellow or red?

Outcome of our interest

A = {Ball drawn is Yellow or Red}

Event

Set of all the possible outcomes

S = {Yellow, Red}

Universal Set

$$P(A) = \frac{\#\text{ of ways A can occur}}{\text{total \# of outcomes}}$$

P(A) = P(red or yellow ball) = 1

Let’s play …

- Assume we throw a die, and note the number it shows
  - What are the possible outcomes now?

Set of all the possible outcomes → S = {1 2 3 4 5 6}

Let’s play …

- What is the probability that it shows 4?

We use this “size” or element counting approach only if all events are **equally likely** (equally likely = have equal chance of happening)

Outcome of our interest $A = \{\text{shows } 4\}$ *Event*

Set of all the possible outcomes $S = \{1\ 2\ 3\ 4\ 5\ 6\}$ *Universal Set*

$$P(A) = \frac{\#\text{ of ways A can occur}}{\text{total }\#\text{ of outcomes}}$$

$$P(A) = 1/6 = \text{size}(A)/\text{size}(S)$$

Let’s play …

- What is the probability that it shows an even number?

Outcome of our interest

A = {shows 2, 4, or 6}

Event

Set of all the possible outcomes

S = {1 2 3 4 5 6}

Universal Set

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total \# of outcomes}}$$

P(A) = 3/6 = size(A)/size(S)

Let's play …

- What is the probability that it shows a number from 1 to 6?

Outcome of our interest

A = {shows 1,2,3,4,5,or 6}

Event

Set of all the possible outcomes

S = {1 2 3 4 5 6}

Universal Set

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total \# of outcomes}}$$

P(A) = 6/6 = size(A)/size(S)

Assigning probabilities is a five-step process

- Step 1: Define the experiment
  - “let’s roll a die”
- Step 2: Define all the possible outcomes of the experiment, i.e., define the universal set that applies to your experiment
  - S = {1 2 3 4 5 6}
- Step 3: Define what outcome you would like to see (we call it “event”)
  - A = {die show even number} = {2 4 6}
- Step 4: Assign probability to the event
  - P(A) = 3/6 = ½

$$P(A) = \frac{\text{\# of ways A can occur}}{\text{total \# of outcomes}}$$

- Step 5: Make sure that the assigned probabilities “make sense”
  - The Three Conditions (discussed later today)!

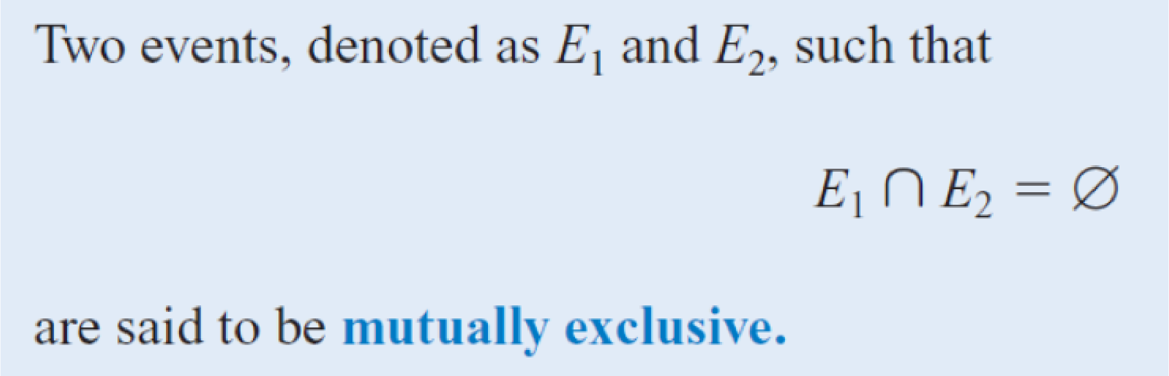
Example Problems

Event Types: **Equally Likely**

- Equally Likely
  - Events that have equal chances of occurring are called “equally likely” events
  - E.g., when tossing a coin, there are equal chances of H and T
  - Sometimes we use common sense to claim that some events are equally likely
  - At other times we may be given equal probabilities assigned to events, and consequently may claim them to be equally likely

Event Types: **Mutually Exclusive**

- Mutually Exclusive
  - Events that cannot occur at the same time are called mutually exclusive
  - E.g., when tossing a coin a head (H) and tail (T) cannot occur at the same time
  - In sets, we check it this way: **if the intersection of two events is empty set then they are mutually exclusive**



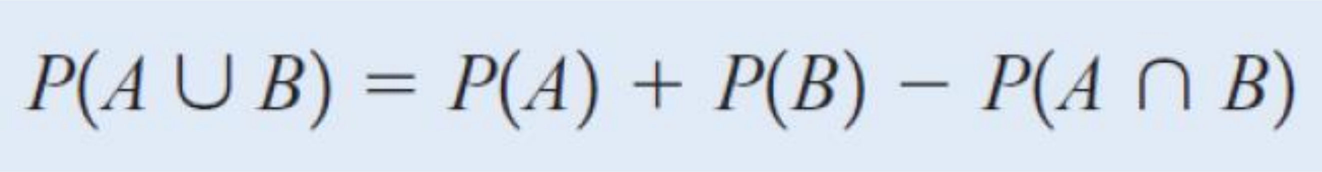
Two events, denoted as E_1 and E_2, such that

$$E_1 \cap E_2 = \varnothing$$

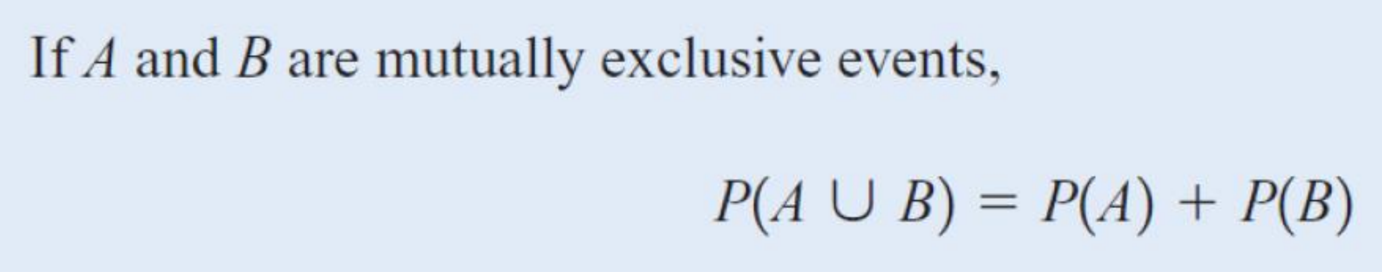
are said to be **mutually exclusive.**

Event Operations: Probability of a Union

P (A ***or*** B)



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



If A and B are mutually exclusive events,

$$P(A \cup B) = P(A) + P(B)$$

Event Operations: Probability of a Union

$$
\begin{aligned}
P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) \\
- P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)
\end{aligned}
$$

Event Operations: Probability of Intersection

P(A ***and*** B) = P(A ∩ B)

- When we assign probabilities to two or more events happening at the same time, we call it **joint probability**
- More on this later…

Event Operations: Probability of Complement

P (***not*** E)

$$P(E') = 1 - P(E)$$

Assigned probabilities must satisfy __three conditions__!!

- Condition 1: Assigned probabilities should not be negative or greater than 1.

$$0 \leq P(E) \leq 1$$

- Condition 2: Universal set should cover all the possible outcomes

$$P(S) = 1$$

- Condition 3: Probabilities assigned to **mutually exclusive** events should make sense

For two events E_1 and E_2 with $E_1 \cap E_2 = \varnothing$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

Let’s check …

S = {1 2 3 4 5 6} *Universal Set*

A = {1}, B = {2}, C ={3}, D = {4}, E = {5}, F = {6}, G = {2 4 6}, … *Events*

P(A) = 1/6 , P(B) = 1/6, P(C) = 1/6, P(D) = 1/6, P(E) = 1/6, P(F) = 1/6, P(G) = 1/2

Assigned probabilities are not negative ✓

P(S) = P (die shows 1, 2, 3, 4, 5, or 6) = 1 *Defined universal set covers all the possible outcomes* ✓

P(A U B) = P(A) + P(B)

P(A U B U C) = P(A) + P(B) + P(C) *Mutually exclusive events satisfy the third condition* ✓

etc…

Example Problems

Previously, we played dice…

We were assigning probabilities to individual events

- We throw a die, and note the number it shows
    - What are the possible outcomes?
- What is the probability that it shows 4?

Outcome of our interest $A = \{\text{shows } 4\}$ *Event*

Set of all the possible outcomes $S = \{1\ 2\ 3\ 4\ 5\ 6\}$ *Universal Set*

$$P(A) = 1/6 = \text{size}(A)/\text{size}(S)$$

Sometimes we need to assign probabilities to multiple events!!

- Let’s say its 8 am, and let’s define two events

A = {Ali is in lecture} *Event one*

B = {Ali is sleeping} *Event two*

- What is the probability that Ali is in lecture **and** sleeping?

P(A **and** B)

Joint Probability

- When we assign probabilities to two or more events happening at the same time, we call it **joint probability**

P(A **and** B)

- Using sets, we can define the joint probability of two events as

P(A **and** B) = $P(A \cap B)$

Let’s play again …

- We throw a die, and note the number it shows
- Let us define two events of interest

A = {shows odd number} *Event one*

B = {shows even number} *Event two*

S = {1 2 3 4 5 6} *Universal Set*

$P(A \textbf{ and } B) = P(A \cap B) = 0$ *How?*

$A \cap B = \{\}$ (empty)

$P(A \cap B) = \text{size}(A \cap B)/\text{size}(S) = 0/6=0$

Let’s play again …

- We throw a die, and note the number it shows

C = {shows 1, 2 or 3} *Event one*

D = {shows 2, 3, or 4} *Event two*

S = {1 2 3 4 5 6} *Universal Set*

P(C **and** D) = P(C ∩ D) = 1/3 *How?*

C ∩ D = {2 3}

P(C ∩ D) = size(C ∩ D)/size(S) = 2/6=1/3

Sometimes knowledge of one random event can help us assign probability to another random event

A = {Ali is in lecture at 8 am} *Event one*

B = {Ali is sleeping at 8 am} *Event two*

- Suppose I tell you that Ali is in the lecture at 8 am. Now what are the chances that he is sleeping?

P(B **given** A)

Let’s toss a coin …

- Suppose I toss a coin but do not show you the result
    - What is the probability that it shows Head?
    - A = {shows Head}
    - S = {H T} (universal set)
    - P(A) = ½
- Suppose now that I tell you that it shows Tail
    - Now what is the probability that it shows Head?

Let’s toss a coin…

Situation **before** additional information

A = {shows Head}

S = {H T} *Universal Set*

P{A} = 1/2

Situation **after** additional information

A = {shows Head}

$S = \{\cancel{H}\ T\}$

“H” is no longer in the set of possibilities (it has become impossible)

P{A| *new information*} = 0

Another example …

- Suppose I write a number from 1 to 4 on a piece of paper but do not show you what I wrote
  - What is the probability that I wrote 3?
    - A = {wrote 3}
    - S = {1 2 3 4} (universal set)
    - P(A) = size(A)/size(S) = ¼
  - What is the probability that I wrote an odd number?
    - B = {wrote 1 or 3}
    - S = {1 2 3 4} (universal set)
    - P(B) = size(B)/size(S) = 2/4 = ½
- Suppose now that I tell you that I wrote an odd number.
  - Now what is the probability that I wrote 3?

Another example …

Situation **before** additional information

A = {wrote 3}

S = {1 2 3 4} *Universal Set*

P{A} = ¼

Situation **after** additional information

A = {wrote 3}

B = {wrote odd number}

$S_{new} = \{1\ \cancel{2}\ 3\ \cancel{4}\}$ *Updated Universal Set*

P{A ***given*** B} = P{A|B} = 1/2

Conditional Probability

- When we assign probabilities to an event assuming (or knowing) that another has happened, we call it **conditional probability**

P(A **given** B) *also written as* $P(A \mid B)$

- Using sets, we **define** the **conditional probability** as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Joint probability of A and B

Important assumption : $P(B) \neq 0$

Let’s play again …

- We throw a die, and note the number it shows

A = {shows 1, 2 or 3}

B = {shows 2, 3, or 4}

S = {1 2 3 4 5 6}

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{1/2} = 2/3$$

How?

$A \cap B = \{2\ 3\}$

$P(A \cap B) = \text{size}(A \cap B)/\text{size}(S) = 2/6=1/3$

$P(B) = \text{size}(B)/\text{size}(S) = 3/6=1/2$

Example Problems

Independence : when knowledge of one event does not change the probability of another

- Sometimes knowledge of one event does not affect the probability of another event.
  - In such a situation, we say that the two events are **statistically independent**
  - Example: knowing what time I woke up this morning will not affect the result of a coin toss
- How can we check independence?

$$P(A \mid B) = P(A)$$

***Test 1**: If the conditional probability remains unchanged, then the two events are independent*

$$P(A \cap B) = P(A)\,P(B)$$

***Test 2**: If the joint probability is simply the product of individual probabilities, then the two events are independent*

When we know events are independent…

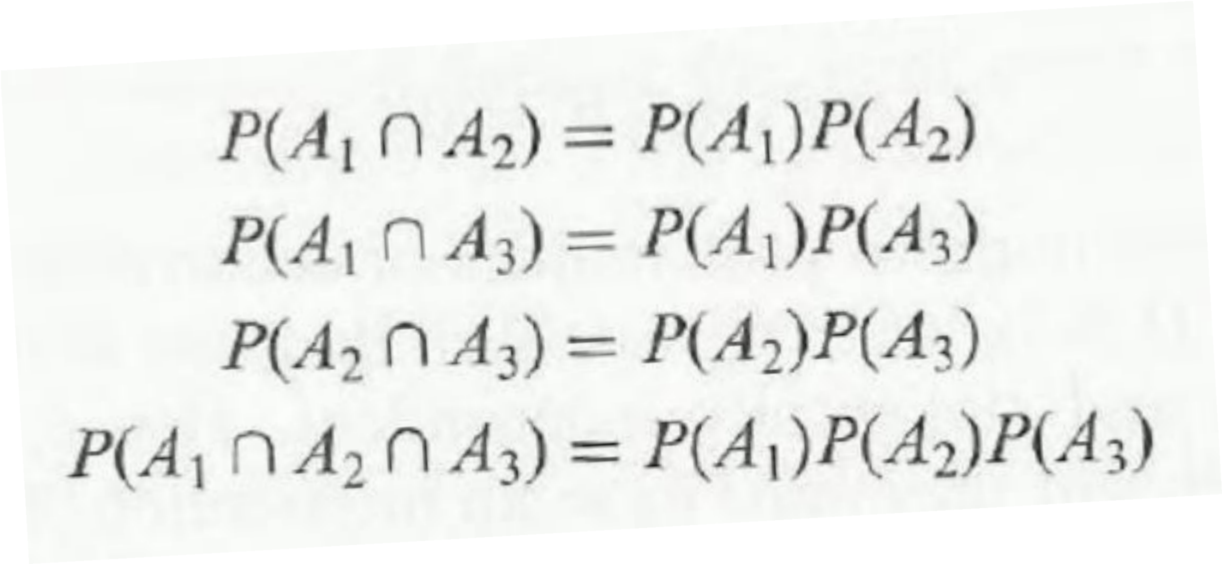
- An important consequence of knowing that two events are independent is that we can use the following substitutions where needed

$$P(A \mid B) = P(A)$$

$$P(A \cap B) = P(A)\,P(B)$$

Can we check independence of more than two events?

- Yes!
    - For **three events to be statistically independent**, all their combinations should satisfy test 2, i.e., they must satisfy **all** of the following



$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

- Similarly, for **N events to be statistically independent**, all their combinations should satisfy test 2.

Example Problems

Two formulas that help us in assigning probabilities

Total Probability

$$P(A) = \sum_{n=1}^{N} P(A|B_n)P(B_n)$$

Bayes' Theorem

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A)}$$

Events $\boldsymbol{B_n}$ *must satisfy these conditions*

$$\bigcup_{n=1}^{N} B_n = S$$

$$B_m \cap B_n = \varnothing$$

i.e., these should be mutually exclusive and should cover all S.

Example Problems

Counting

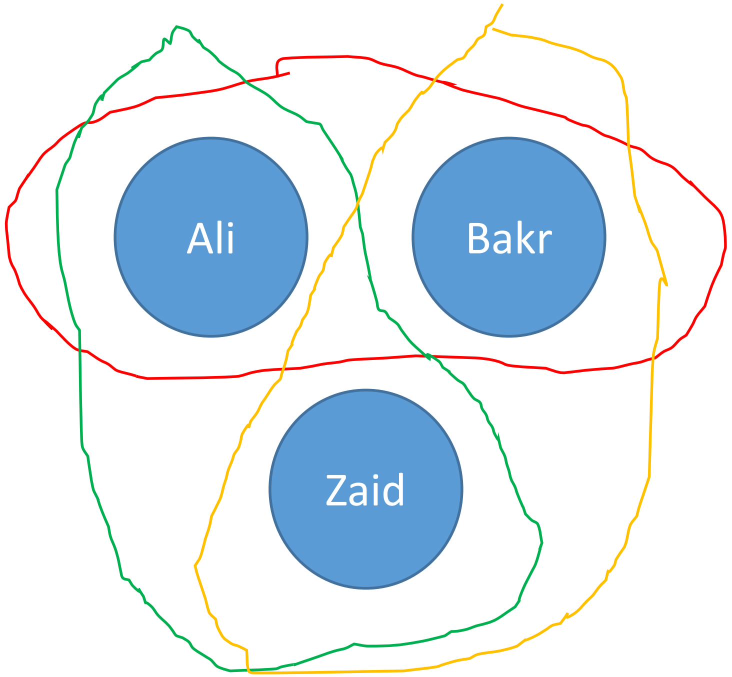
- In assigning probabilities we often need to "count" how many ways something can happen
  - Toss a coin twice, and define event
  - A = {we get at least one Tails}
  - S = {HH HT TH TT}
  - **Count** total number of ways two coin tosses can result: **four** (HH, HT, TH, TT)
  - **Count** number of ways A happens: **three** (HT , TH, TT)
  - P(A) = 3/4

Counting: Two Questions to Ask

- Question1: Does order matter?
- Question2: Is repetition (replacement) allowed?

Question1: Does order matter?
Question2: Is repetition allowed?

Game: choose two of these and I will give 10 Riyals to each. Cannot choose same person twice.



Here:

1. ORDER DOESN'T MATTER
2. REPETITION NOT ALLOWED

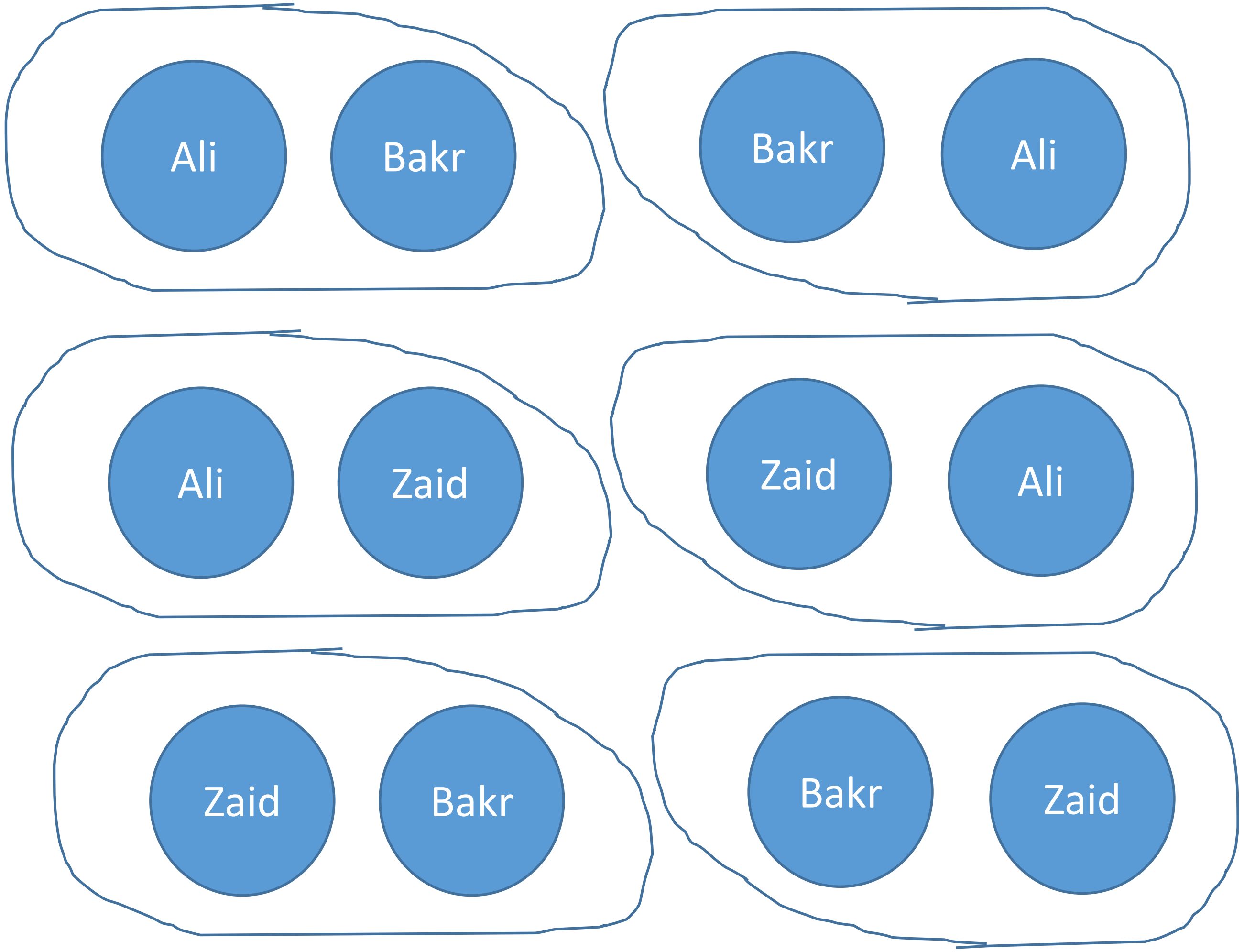
Possible choices: **three**

Question1: Does order matter?
Question2: Is repetition allowed?

Game: choose two of these. I will give 10 Riyals to first one and 5 Riyals to second one. Cannot choose same person twice.

Here:
(1) ORDER MATTERS
(2) REPETITION NOT ALLOWED

Possible choices: **six**



Question1: Does order matter?
Question2: Is repetition allowed?

Game: choose two of these. I will give 10 Riyals to first one and 5 Riyals to second one. You are allowed to choose the same person twice.

Here:
(1) ORDER MATTERS
(2) REPETITION IS ALLOWED

Possible choices: **nine**

Ali Bakr	Bakr Ali	Ali Ali
Ali Zaid	Zaid Ali	Bakr Bakr
Zaid Bakr	Bakr Zaid	Zaid Zaid

Note: most number of possibilities when order matters and repetition is allowed!

How to choose counting formula!

Recall **5! = 5 x 4 x 3 x 2 x 1**
And **0! = 1**

Total objects = **n**
Objects to choose (or number of places to fill) = **r**

Note: for "without replacement" cases
$r \leq n$

Order Matters?	Repetition Allowed?	What to do.	Formula
Yes	Yes	Use Powers	n^r
Yes	No	Use Permutations	${}^nP_r = \frac{n!}{(n-r)!}$
No	No	Use Combinations	${}^nC_r = \frac{n!}{(n-r)!\,r!}$

Applying the counting formulas

Ali

Bakr

Zaid

Game: choose two of these three students and I will give 10 Riyals to each. Cannot choose same person twice.

Here:

(1) ORDER DOESN'T MATTER
(2) REPETITION NOT ALLOWED
(3) $\mathbf{n = 3, r = 2}$

$$^{n}C_{r} = \frac{n!}{(n-r)!\,r!}$$

$$^{3}C_{2} = \frac{3!}{(3-2)!\,2!} = 3$$

Game: choose two of these three students and I will give 10 Riyals to first one and 5 Riyals to second one. Cannot choose same person twice.

Here:

(1) ORDER MATTERS
(2) REPETITION NOT ALLOWED
(3) $\mathbf{n = 3, r = 2}$

$$^{n}P_{r} = \frac{n!}{(n-r)!}$$

$$^{3}P_{2} = \frac{3!}{(3-2)!} = 3! = 6$$

Game: choose two of these three students and I will give 10 Riyals to first one and 5 Riyals to second one. You are allowed to choose the same person twice.

Here:

(1) ORDER MATTERS
(2) REPETITION ALLOWED
(3) $\mathbf{n = 3, r = 2}$

$$n^{r} = 3^{2} = 9$$

Example Problems

Questions?? Thoughts??

EE 302
Probabilistic Methods in Electrical Engineering

with

Dr. Naveed R. Butt

@

Jouf University

Random Event → Random Variable

- Let’s talk about the next student who enters the room.
  - X = number of mobiles he has
  - Y = his height
  - Z = is he from Jouf region?
- What are the possible values of X, Y, and Z?

Random Event → Random Variable

- Let’s talk about the next student who enters the room.
  - X = number of mobiles he has
  - Y = his height
  - Z = is he from Jouf region?
- What are the possible values of X, Y, and Z?
  - X = 0, 1, 2, 3 … mobiles. Only fixed values (**discrete**)
  - Y = Height in cm. Any value in a certain range (**continuous**)
  - Z = Yes or No (just description, **no numeric values**)

Random Event → Random Variable

- All three, X, Y, and Z are random events, but …
  - X is a **discrete** random variable
  - Y is a **continuous** random variable
  - Z is **not a random variable**! (to qualify for random variable it must have numeric values)

What is a Random Variable?

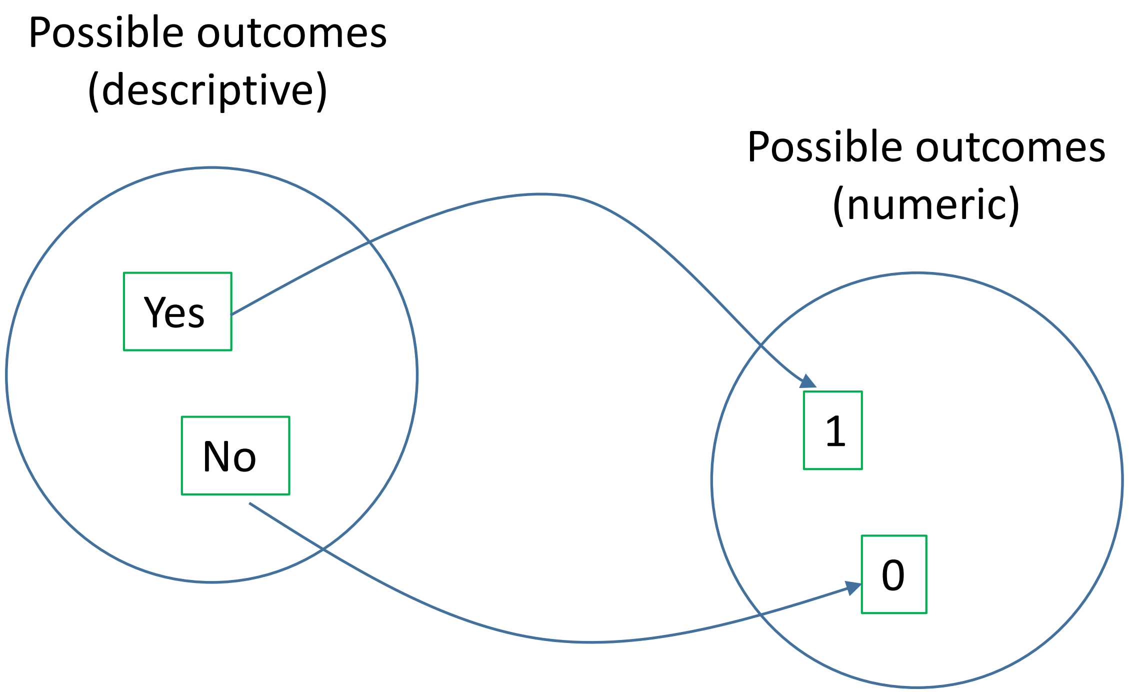
- A random variable
  1. Is a variable whose value depends on the result of a random occurrence
  2. Always takes numeric values (NOT descriptions)
  3. Can be continuous or discrete
  4. Must satisfy some additional conditions (such as $P(X = \infty) = 0$ etc.)

Random Event → Random Variable

- We can convert a random event to random variable by assigning numeric values to descriptive outcomes

Z = Next student who enters the room, is he from Jouf region?

Not a random variable yet.



Possible outcomes
(descriptive)

Possible outcomes
(numeric)

Yes

No

1

0

Z = 1 if next student entering the room is from Jouf
Z = 0 if next student entering the room in not from Jouf

Now Z is a random variable!!

Random Event → Random Variable

- We can convert a random event to random variable by assigning numeric values to descriptive outcomes
- Z = Next student who enters the room, is he from Jouf region?
  - Possible outcomes, Z = Yes or No
  - Z is not a random variable as its values are not numeric
  - We can convert Z into a random variable by mapping Yes = 1, No = 0
  - Then possible outcomes are Z = 1 or 0
    - Now Z is a random variable

Collecting Probabilities

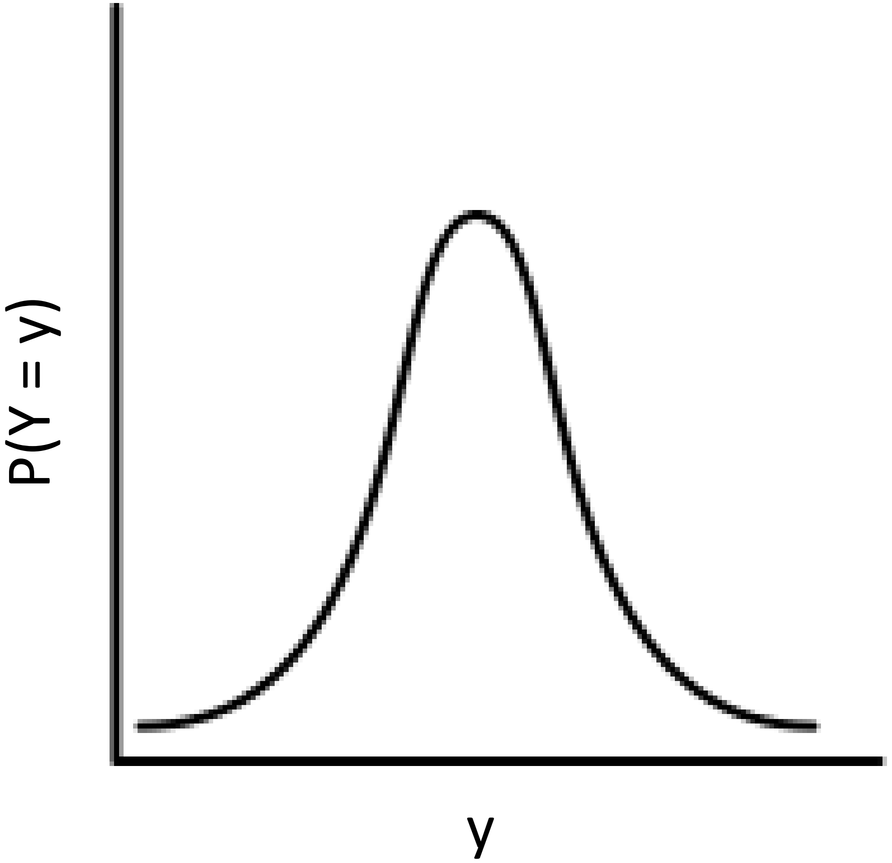
- Once again, let’s talk about the next student who enters the room.
  - X = number of mobiles he has
  - Y = his height
  - X and Y are random variables. We will denote the values they take by small letters x and y.
- How do we represent the probabilities of the different values of X and Y?

We could make a table ...

x	P(X = x)
0	0.1
1	0.6
2	0.2
…..	…..

X = number of mobiles next student entering the room has

Or, we could draw a graph …



Y = height of the next student entering the room

Or, we could write probabilities as a function (formula) …

$$f(z) = \frac{1}{6} \qquad z = 1, 2, \dots, 6$$

Z = result of rolling a die

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu - x)^2}{2\sigma^2}}$$

X = student height

Collecting Probabilities

- How do we represent the probabilities of the different values of X and Y?
  - We could write them in a table
  - Draw them as a graph
  - Or, write them as a mathematical function (formula)

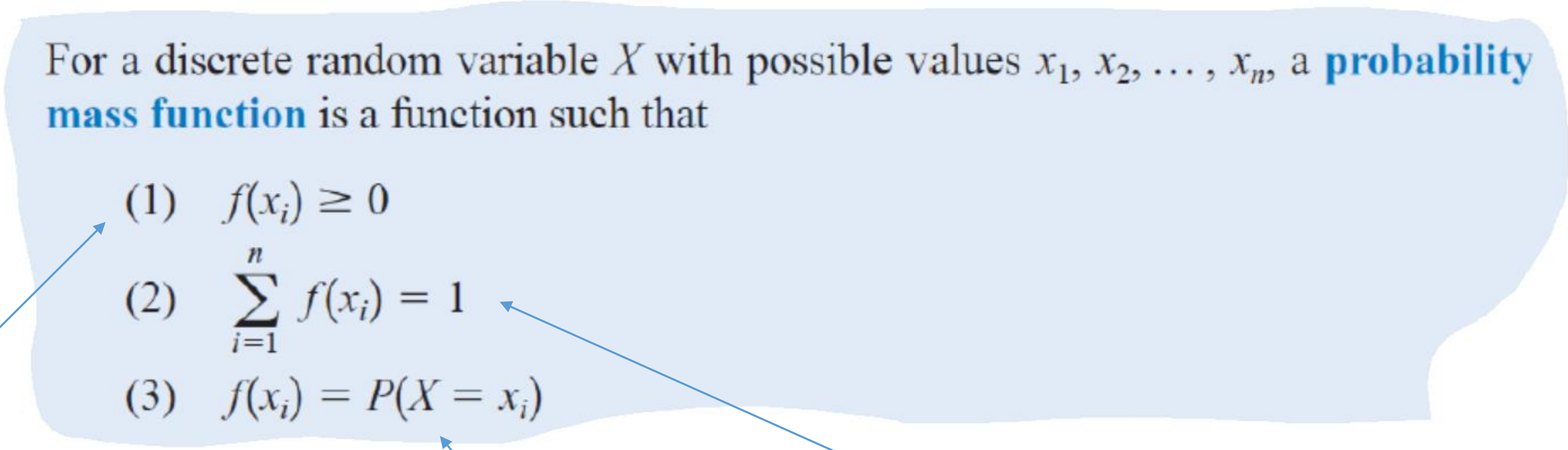
Probability Mass Function (PMF): $f(x)$

- First we will focus on **discrete random variables** (e.g. X in previous slides)
- The formula showing the probabilities of different values of X is called **probability mass function** (PMF)

$$f(x) = P(X = x)$$

Probability that X
takes the value x

Three Conditions $f(x)$ Must Satisfy



For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$

Assigned probabilities cannot be negative

Probabilities of all possible values of X must add up to 1 (neither more nor less than 1)

f(x) must represent probabilities

Let's play …

- Let's assume there is a bag with three balls in it with numbers 1 – 3 written on them. You draw one ball at random.
  - X = number written on the ball (random variable)
  - x = 1, 2, 3 (possible values of X)
  - $f(x) = P(X = x)$

Note that $f(x)$ satisfies the three conditions

1. Its always non-negative
2. Sums up to 1
3. It assigns probabilities to each value of X

$\boldsymbol{x}$	$\boldsymbol{f(x)}$
1	$\frac{1}{3}$
2	$\frac{1}{3}$
3	$\frac{1}{3}$

Cumulative Distribution Function (CDF): $F(x)$

- Sometimes we like to talk about what range of values a random variable may take
    - What is the probability that the next student entering the room has more than two mobiles? i.e., $P(X > 2) = ?$
    - What is the probability that the next student entering the room has two mobiles or less?, i.e., $P(X \leq 2) =?$
- For such cases, the Cumulative Distribution Function (CDF) is useful.

Cumulative Distribution Function (CDF): $F(x)$

$$F(x) = P(X \le x)$$

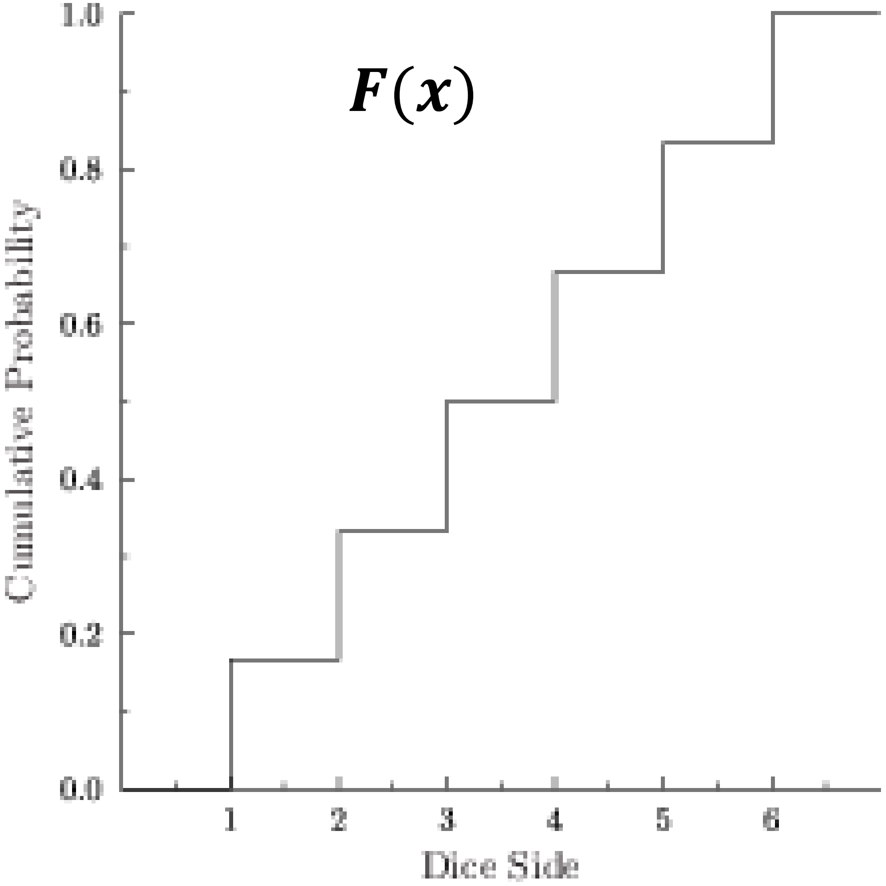
Probability that X takes a value less than or equal to x

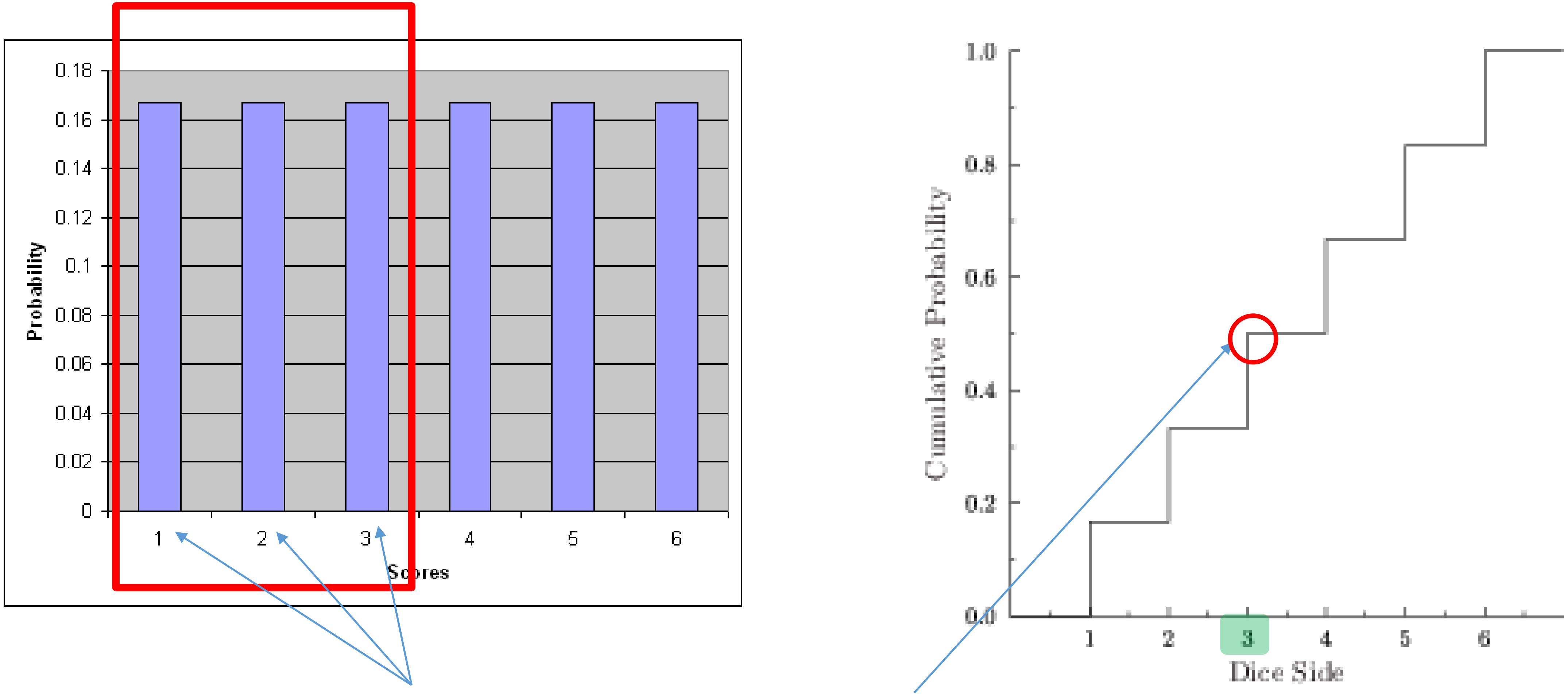
$$F(x) = \sum_{x_i \le x} f(x_i)$$

CDF of a discrete random variable is just the **sum of the PMF** values for $x_i \le x$

Cumulative Distribution Function (CDF): $F(x)$

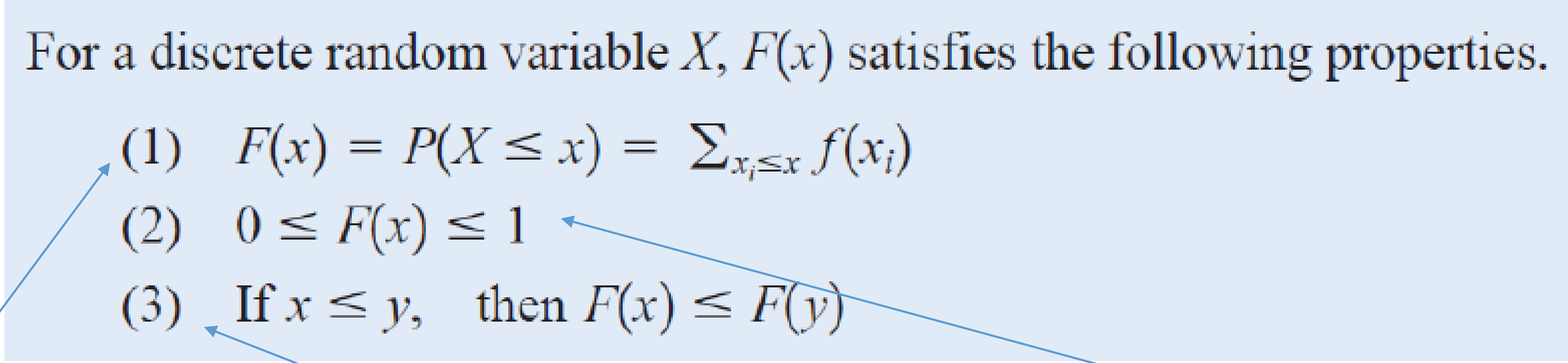
$f(x)$





$P(X \leq 3)$ = sum of all these possibilities = $\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=0.5$

Three Conditions $F(x)$ Must Satisfy



For a discrete random variable X, $F(x)$ satisfies the following properties.

(1) $F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$

(2) $0 \leq F(x) \leq 1$

(3) If $x \leq y$, then $F(x) \leq F(y)$

Like probabilities, CDF always lies **between 0 and 1.**

CDF is always an **increasing function**

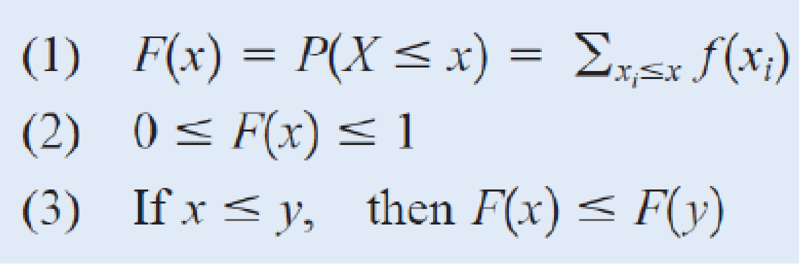
CDF of a discrete random variable is just the **sum of the PMF** values for $x_i \leq x$

Let’s play …

- Let’s write $F(x)$ for the three balls example

$\boldsymbol{x}$	$\boldsymbol{f(x)}$	$F(x) = P(X \le x)$
1	$\frac{1}{3}$	$F(1) = f(1) = \frac{1}{3}$
2	$\frac{1}{3}$	$F(2) = f(1) + f(2) = \frac{2}{3}$
3	$\frac{1}{3}$	$F(3) = f(1) + f(2) + f(3) = 1$

Note that $F(x)$ satisfies the three conditions



(1) $F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$

(2) $0 \le F(x) \le 1$

(3) If $x \le y$, then $F(x) \le F(y)$

We can use $F(x)$ to find many kinds of probabilities (ranges and values)

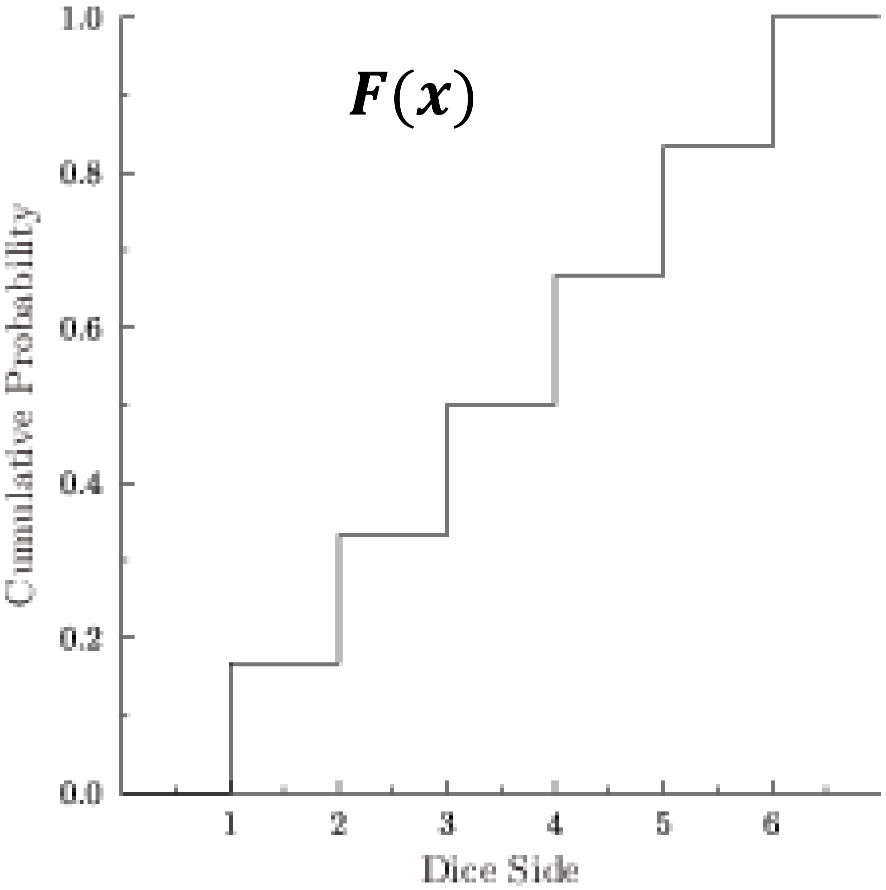
- $P(X\leq a)= F(a)$
- $P(X= a)=P(X\leq a) - P\ (X< a)$

  $=P(X\leq a)- P(X\leq a-1)$

  $=1- F(a) - F\ (a-1)$
- $P(X>a)=1-P(X\leq a)=1 - F(a)$
- $P(X\geq a)=1- P(X<a)=1- P(X\leq a -1)= 1- F(a-1)$
- $P(X<a)=P(X\leq a - 1)=F\ (a -1)$
- $\mathrm{P}(\mathrm{a} <\mathrm{X} \leq \mathrm{b})= \mathrm{F}(\mathrm{b}) - \mathrm{F}(\mathrm{a})$

Test the relations on previous slides on this example

$f(x)$



Joint Probability Distribution (JPD): $f(x, y)$

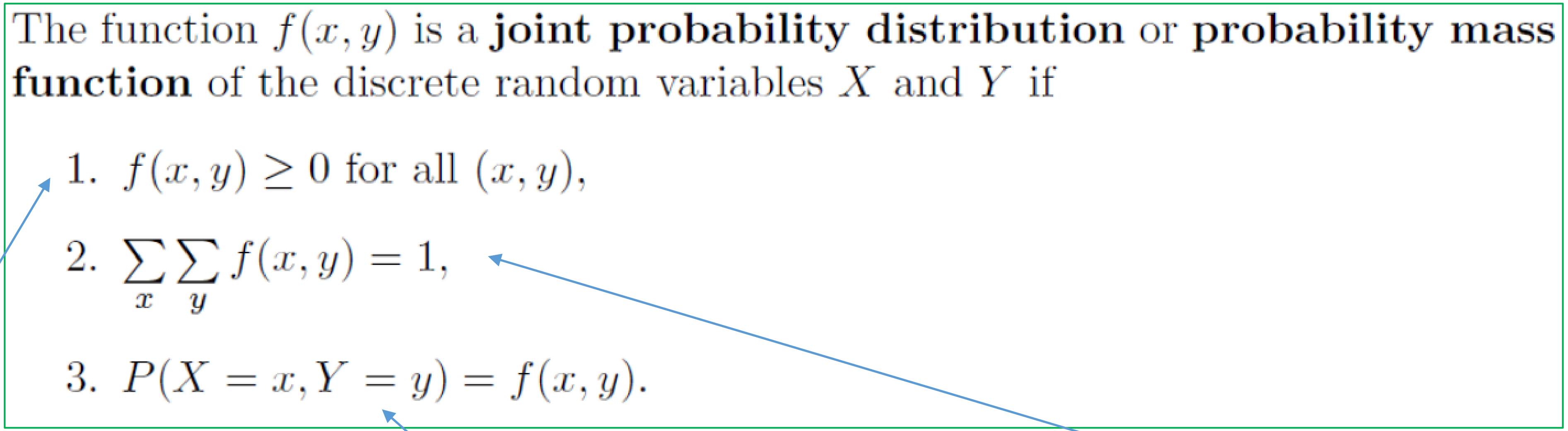
- Sometimes we are interested in finding the probability of two things happening at the same time
  - What is the probability that next student entering the room has two mobiles and no pen?
    - X = number of mobiles he has
    - Y = number of pens he has
    - P(X = 2 **and** Y = 0) = ?
- For such cases, we use the Joint Probability Distribution (JPD)

Joint Probability Distribution (JPD): $f(x,y)$

$$f(x,y) = P(X = x, Y = y)$$

Probability that X takes the value x
and Y takes the value y

Three Conditions $f(x, y)$ Must Satisfy



The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_{x} \sum_{y} f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

JPD **can never be negative**.

JPD **should represent the joint probability** of X and Y.

The **total JPD of two variables should exactly be 1** (neither more nor less).

Let’s play …

- Let’s assume there is a bag with three balls in it with numbers 1 – 3 written on them. You draw two balls (one by one) at random.
  - X = number written on the first ball (random variable)
  - Y = number written on the second ball (random variable)
  - x = 1, 2, 3 (possible values of X)
  - y = 1, 2, 3 (possible values of Y)
  - $g(x) = P(X = x)$
  - $h(y) = P(Y = y)$
  - $f(x, y) = P(X = x, Y = y)$

S = { (1,2) (1,3) (2,1) (2,3) (3,1) (3,3) }

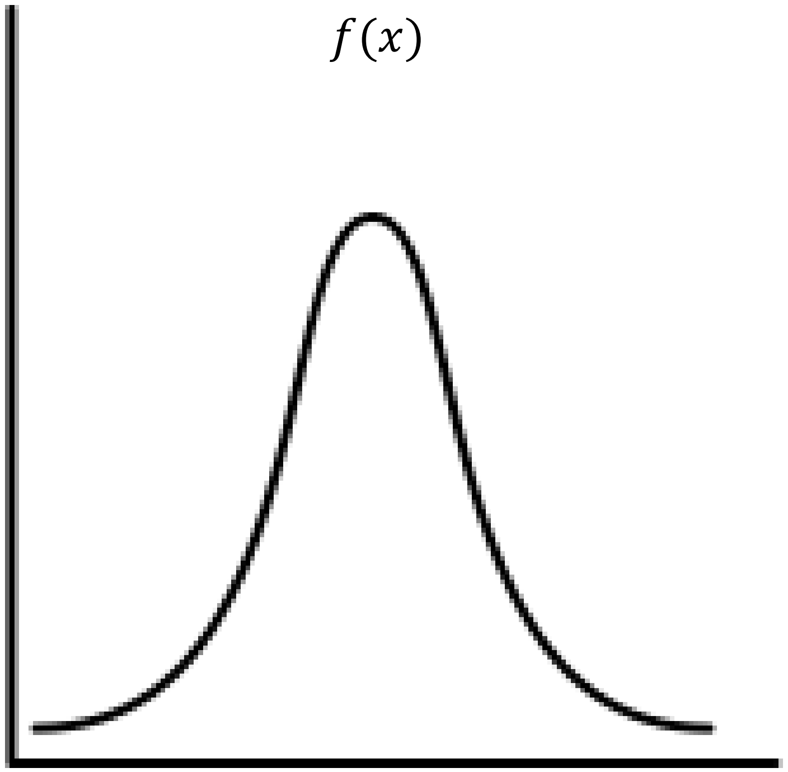
All the possible outcomes (note that they are **equally likely**)

x, y	1	2	3
1	$f(1,1)$	$f(1,2)$	$f(1,3)$
2	$f(2,1)$	$f(2,2)$	$f(2,3)$
3	$f(3,1)$	$f(3,2)$	$f(3,3)$

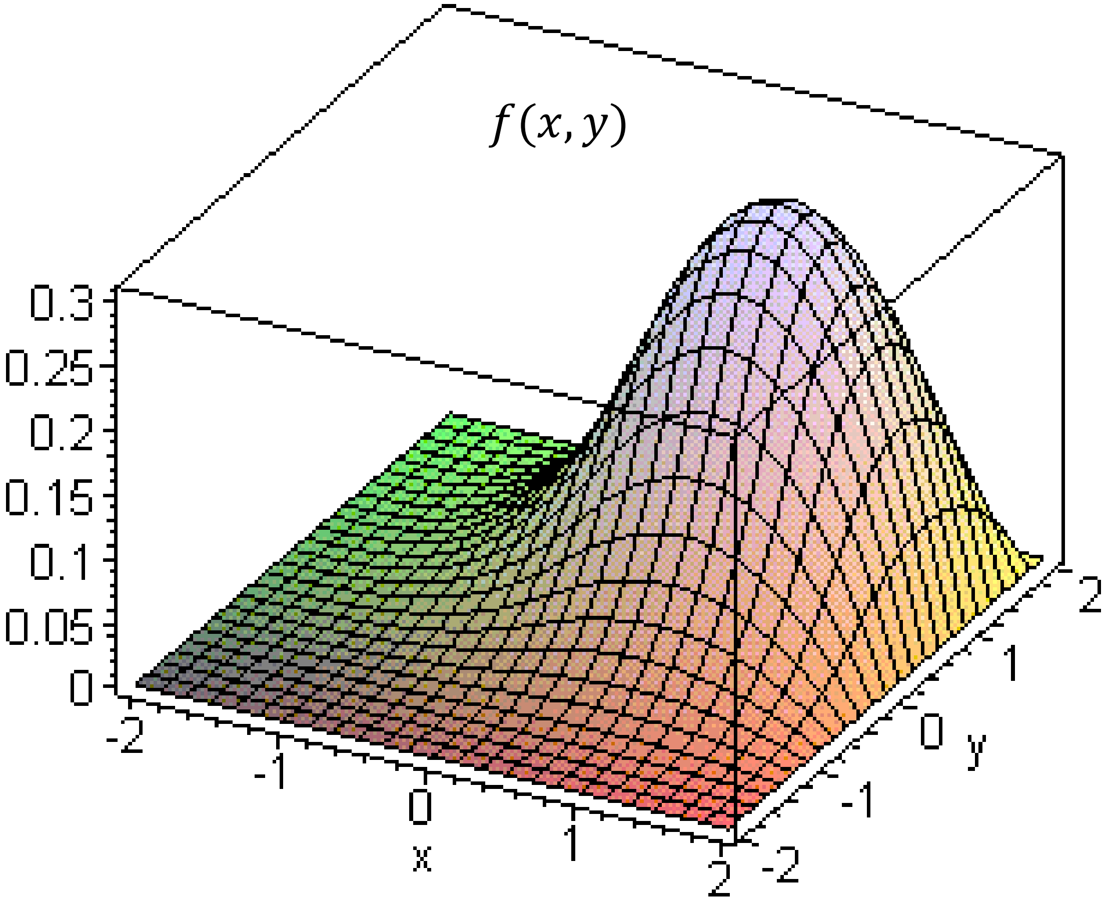
x, y	1	2	3
1	0	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	0	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{1}{6}$	0

Note that $f(x, y)$ satisfies the three conditions

1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.



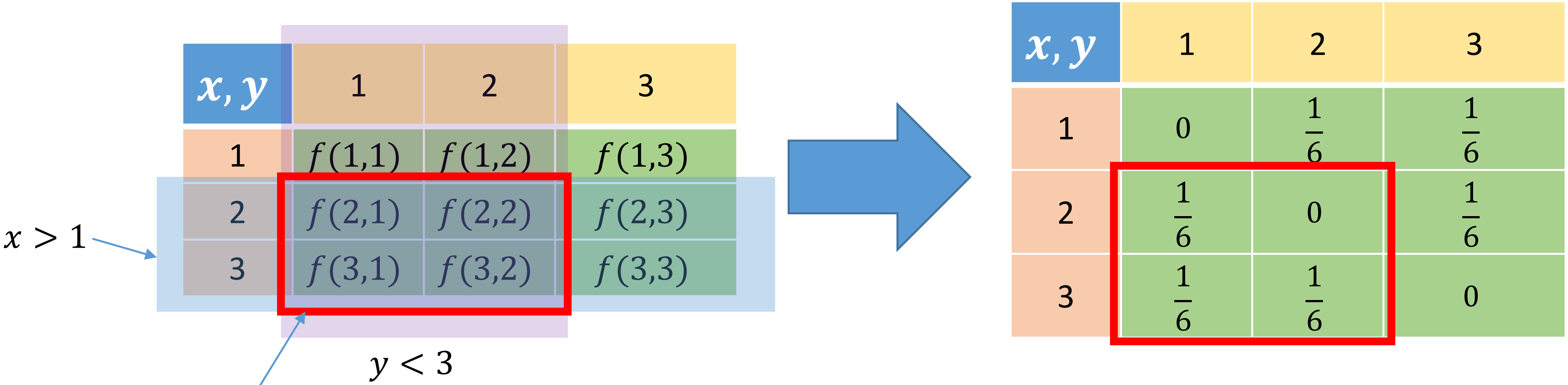
Plot of a single random variable distribution is a two dimensional curve



Plot of a joint distribution is a three dimensional curve!

Probability over an area *A*

For any region A in the xy plane, $P[(X,Y) \in A] = \sum\sum_{A} f(x,y)$.



x, y	1	2	3
1	$f(1,1)$	$f(1,2)$	$f(1,3)$
2	$f(2,1)$	$f(2,2)$	$f(2,3)$
3	$f(3,1)$	$f(3,2)$	$f(3,3)$

$x > 1$

$y < 3$

x, y	1	2	3
1	0	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	0	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{1}{6}$	0

$A = \{x > 1, y < 3\}$

$$P[(X,Y) \in A] = f(2,1) + f(2,2) + f(3,1) + f(3,2) = \frac{1}{2}$$

Marginal Distribution

- Sometimes we are interested in questions such as “given the joint distribution of X and Y, find the distribution of X”
  - This may be the case, for instance, if we have easy access to the joint distribution.
- The distribution of X extracted from the joint distribution of X and Y (by summing over all possible values of Y) is called the **Marginal Distribution of X**

Marginal Distribution

Sum $f(x, y)$ over all the possible values of Y

Marginal Distribution of X

$$g(x) = \sum_{y} f(x, y)$$

Marginal Distribution of Y

$$h(y) = \sum_{x} f(x, y)$$

Sum $f(x, y)$ over all the possible values of X

x, y	1	2	3
1	$f(1,1)$	$f(1,2)$	$f(1,3)$
2	$f(2,1)$	$f(2,2)$	$f(2,3)$
3	$f(3,1)$	$f(3,2)$	$f(3,3)$

$g(1) = f(1,1) + f(1,2) + f(1,3)$

$g(2) = f(2,1) + f(2,2) + f(2,3)$

$g(3) = f(3,1) + f(3,2) + f(3,3)$

$g(x)$ = marginal distribution of *x*

x, y	1	2	3
1	$f(1,1)$	$f(1,2)$	$f(1,3)$
2	$f(2,1)$	$f(2,2)$	$f(2,3)$
3	$f(3,1)$	$f(3,2)$	$f(3,3)$

$h(1) = f(1,1) + f(2,1) + f(2,3)$

$h(2) = f(1,2) + f(2,2) + f(3,2)$

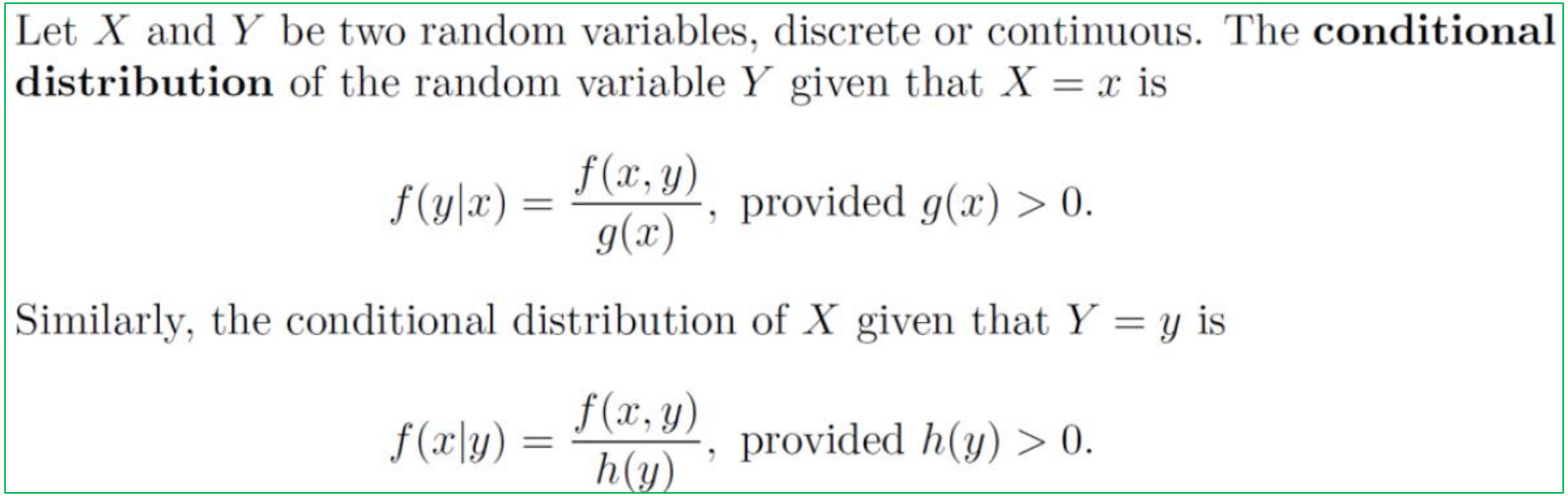
$h(3) = f(1,3) + f(2,3) + f(3,3)$

$h(y)$ = marginal distribution of *y*

Conditional Distribution (CD): $f(x \mid y)$

- Sometimes we are interested in finding the **probability of an event given that another event has taken place**.
  - What is the probability that a student will get GPA 5 in this semester given that his CGPA till previous semester is 4.5?
    - X = GPA student will get in this semester
    - Y = His CGPA now
    - P(X = 5 **given** Y = 4.5) = ?
- For such cases, we use the Conditional Distribution (CD)

Conditional Distribution (CD): $f(x \mid y)$



Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \text{ provided } h(y) > 0.$$

Note:
$g(x)$ = marginal distribution of X
$h(y)$ = marginal distribution of Y

S = { (1,2) (1,3) (2,1) (2,3) (3,1) (3,3) } ← All the possible outcomes (note that they are **equally likely**)

x

y

$y\|x$	1	2	3
1	$f(1\|1)$	$f(1\|2)$	$f(1\|3)$
2	$f(2\|1)$	$f(2\|2)$	$f(2\|3)$
3	$f(3\|1)$	$f(3\|2)$	$f(3\|3)$

$y\|x$	1	2	3
1	0	$\frac{1}{2}$	$\frac{1}{2}$
2	$\frac{1}{2}$	0	$\frac{1}{2}$
3	$\frac{1}{2}$	$\frac{1}{2}$	0

Statistical Independence

- Two random variables are statistically independent if **knowledge of one does not change the probability distribution of the other**.

How to check statistical independence?

- We can prove that two random variables are statistically independent by showing any of the following to be **true for all values of *x* and *y***.

$$\boxed{f(x|y) = g(x)}$$

$$\boxed{f(y|x) = h(y)}$$

$$\boxed{f(x,y) = g(x)h(y)}$$

Note:
$g(x)$ = marginal distribution of X
$h(y)$ = marginal distribution of Y

S = { (1,2) (1,3) (2,1) (2,3) (3,1) (3,3) } ←

All the possible outcomes (note that they are **equally likely**)

x, y	1	2	3
1	$f(1,1)$	$f(1,2)$	$f(1,3)$
2	$f(2,1)$	$f(2,2)$	$f(2,3)$
3	$f(3,1)$	$f(3,2)$	$f(3,3)$

x, y	1	2	3
1	0	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	0	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{1}{6}$	0

X and Y are not statistically independent, since for many values of x and y

$$f(x, y) \neq g(x)h(y)$$

e.g., $f(2,1) \neq g(2)h(1)$

Questions?? Thoughts??

EE 302
Probabilistic Methods in Electrical Engineering

with

Dr. Naveed R. Butt

@

Jouf University

We have previously talked about …

Collecting Probabilities

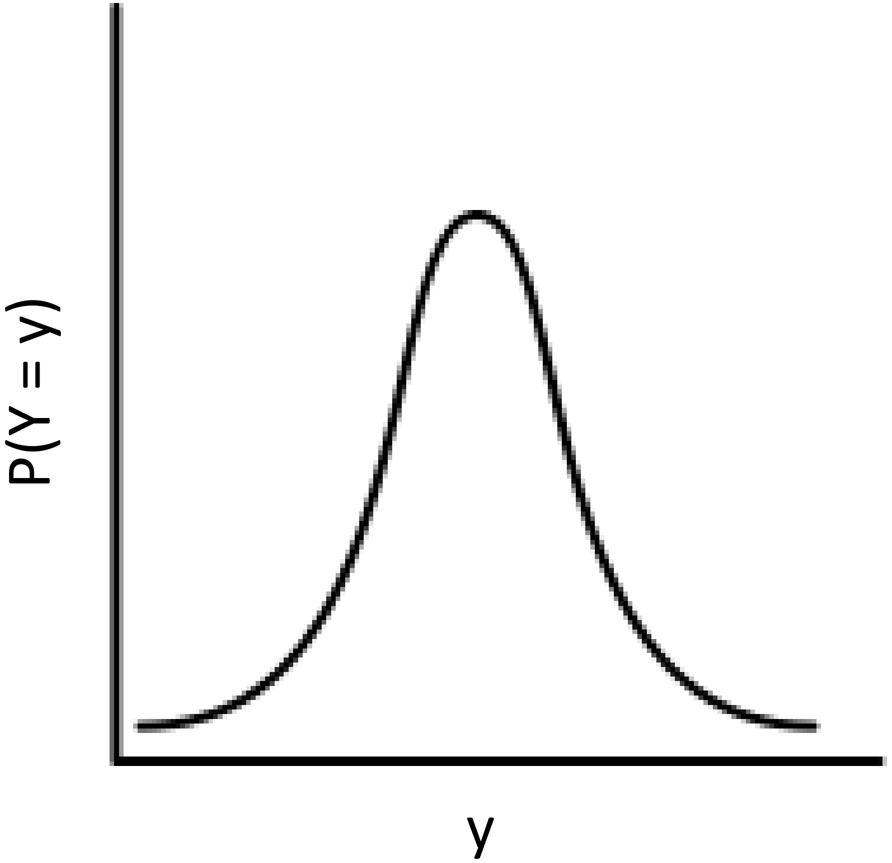
- Next student who enters the room.
  - X = number of mobiles he has
  - Y = his height
- **How do we represent the probabilities of the different values of X and Y?**

We could make a table …

x	P(X = x)
0	0.1
1	0.6
2	0.2
…..	…..

X = number of mobiles next student entering the room has

Or, we could draw a graph …



Y = height of the next student entering the room

Or, we could write probabilities as a function (formula) …

$$f(z) = \frac{1}{6} \qquad z = 1, 2, \dots, 6$$

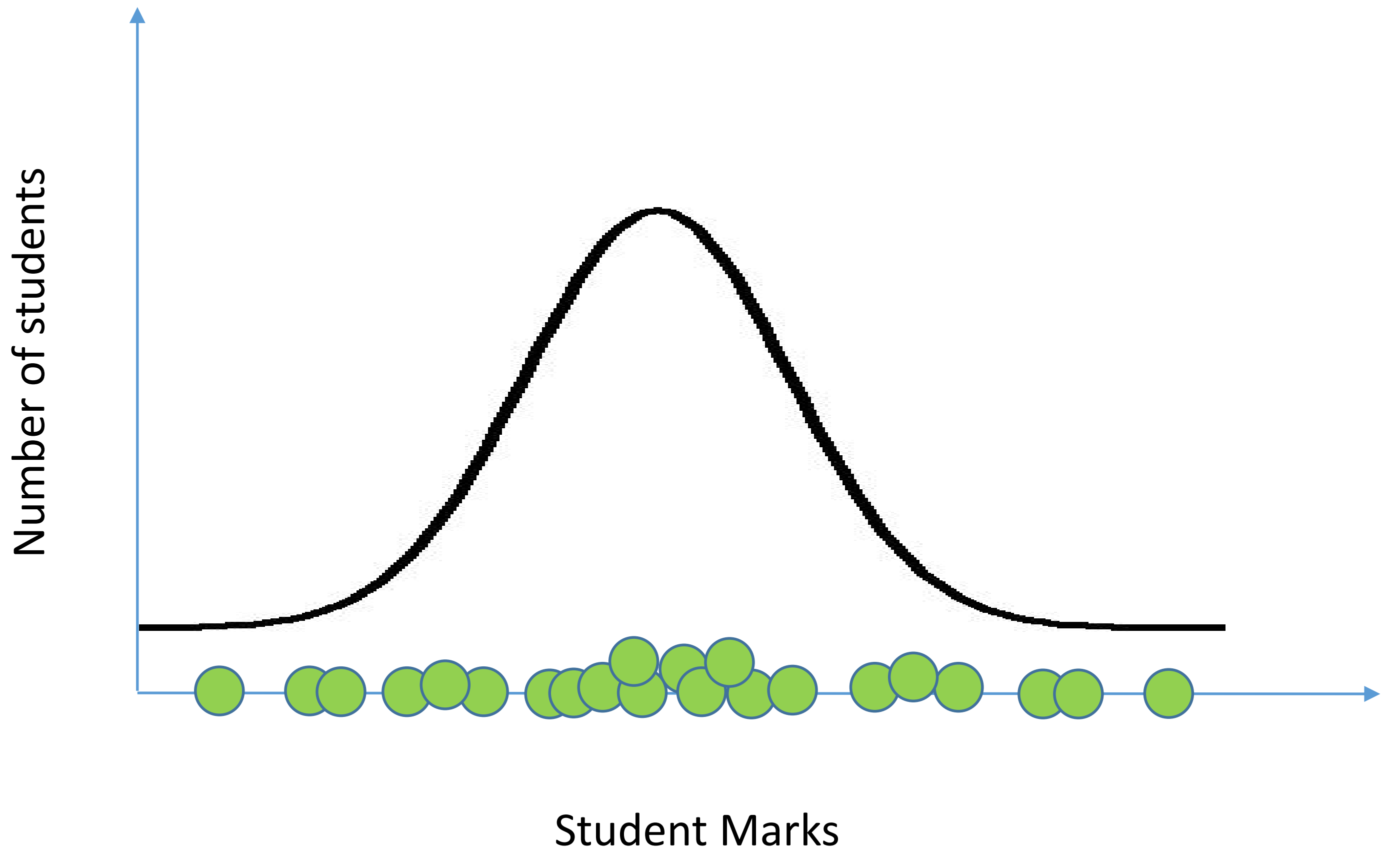
Z = result of rolling a die

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mu - x)^2}{2\sigma^2}}$$

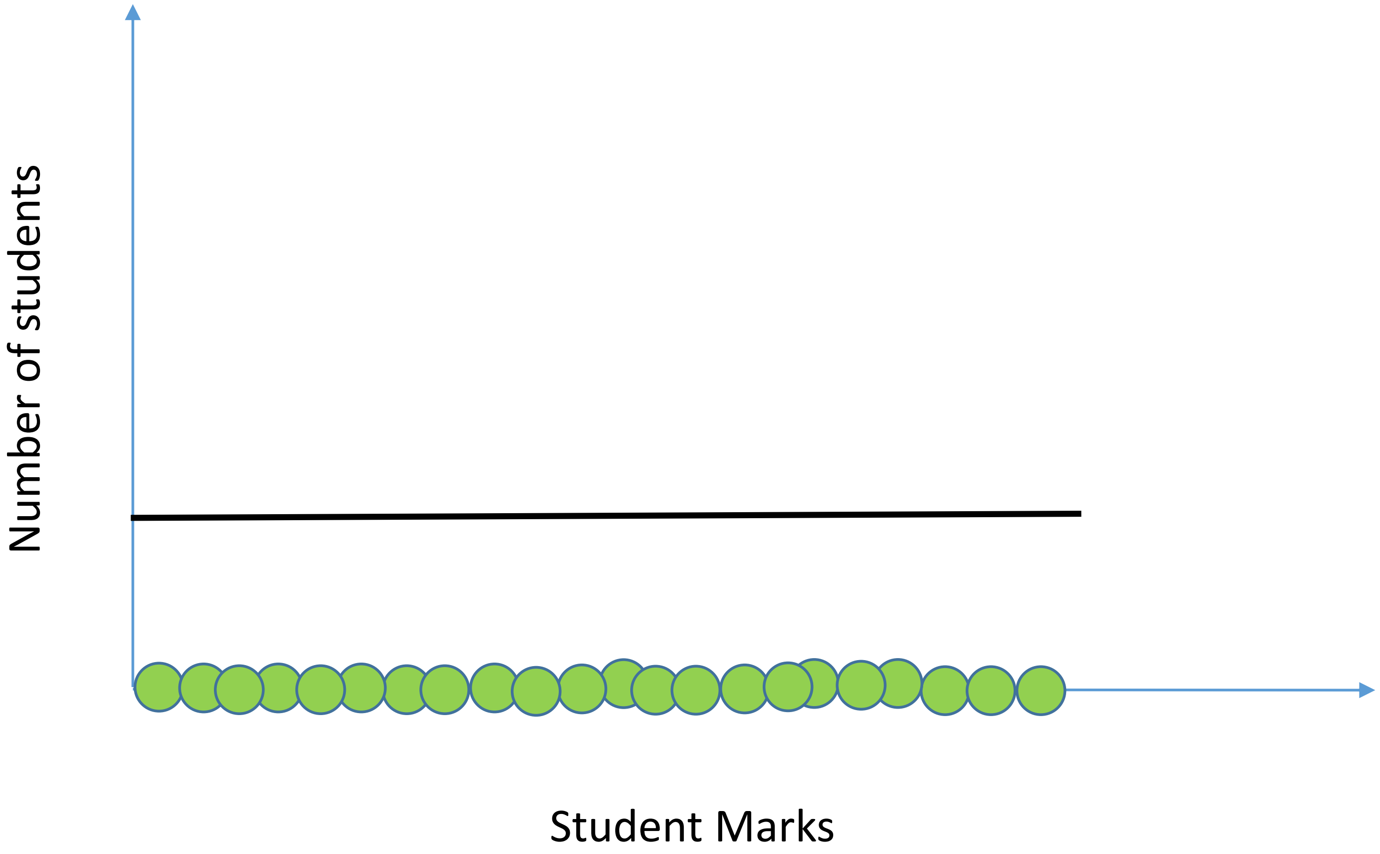
X = student height

Collecting Probabilities

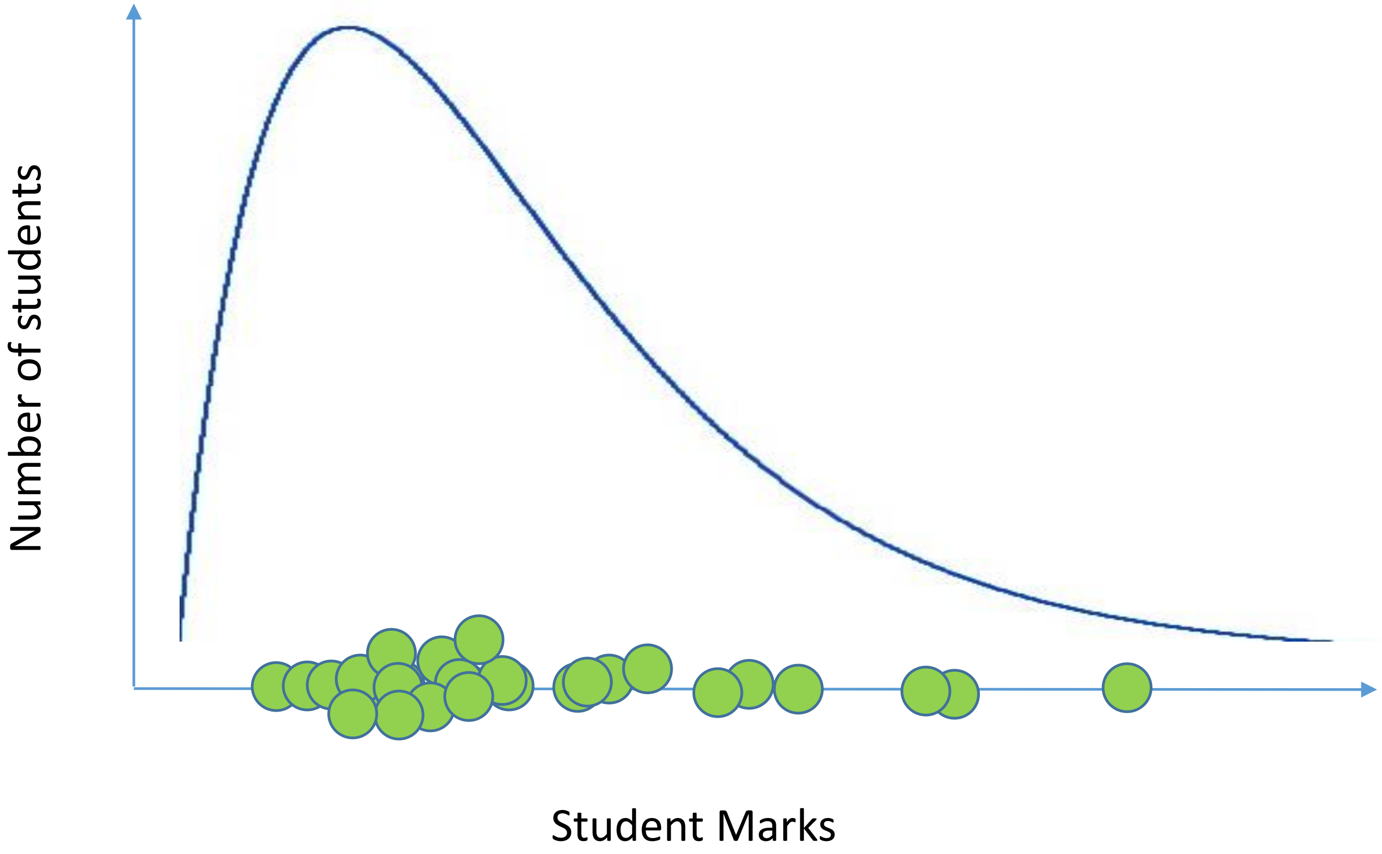
- So, we collect probabilities in terms of a probability mass function (discrete case) or probability density function (continuous case)



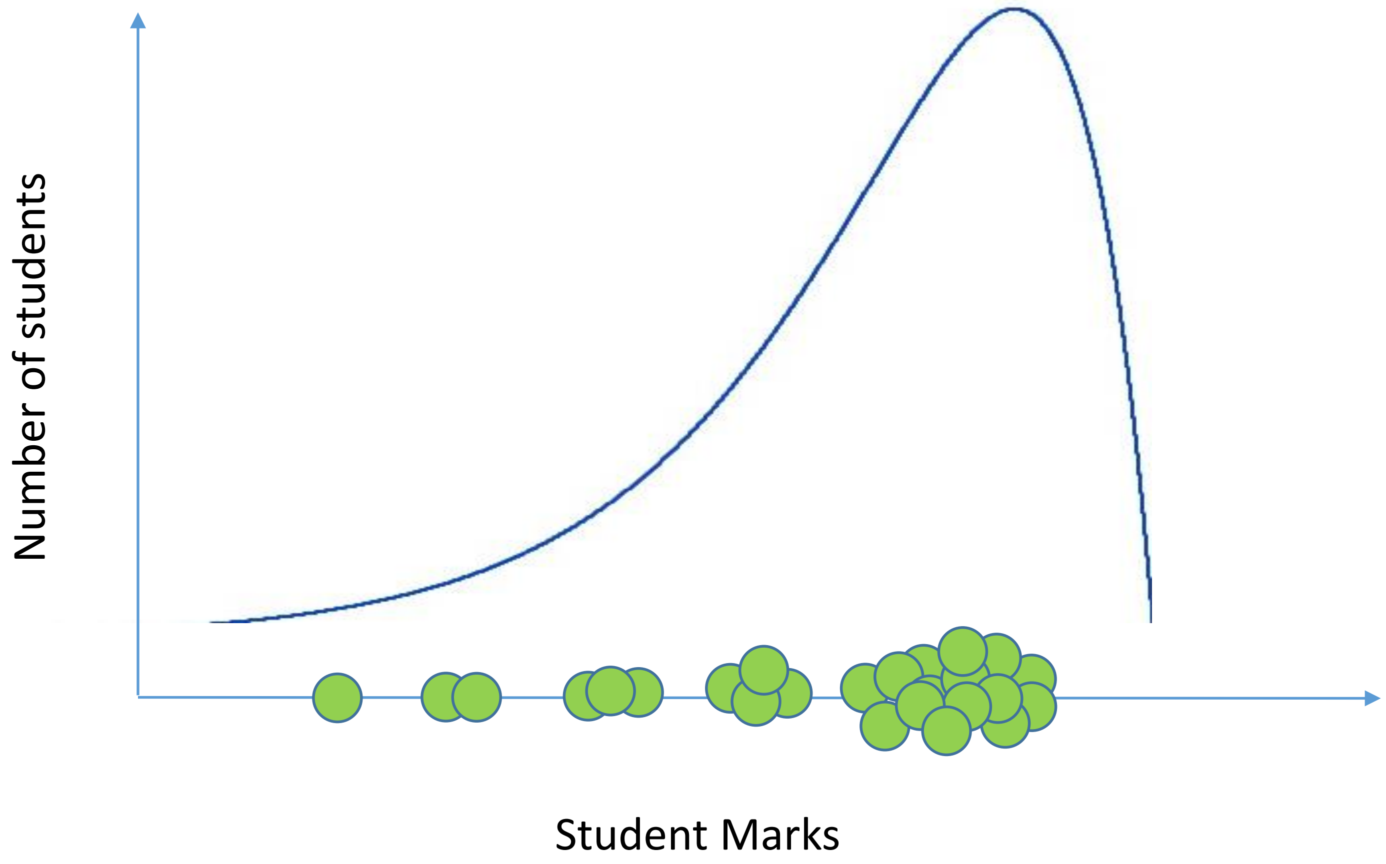
Collecting Probabilities



Collecting Probabilities



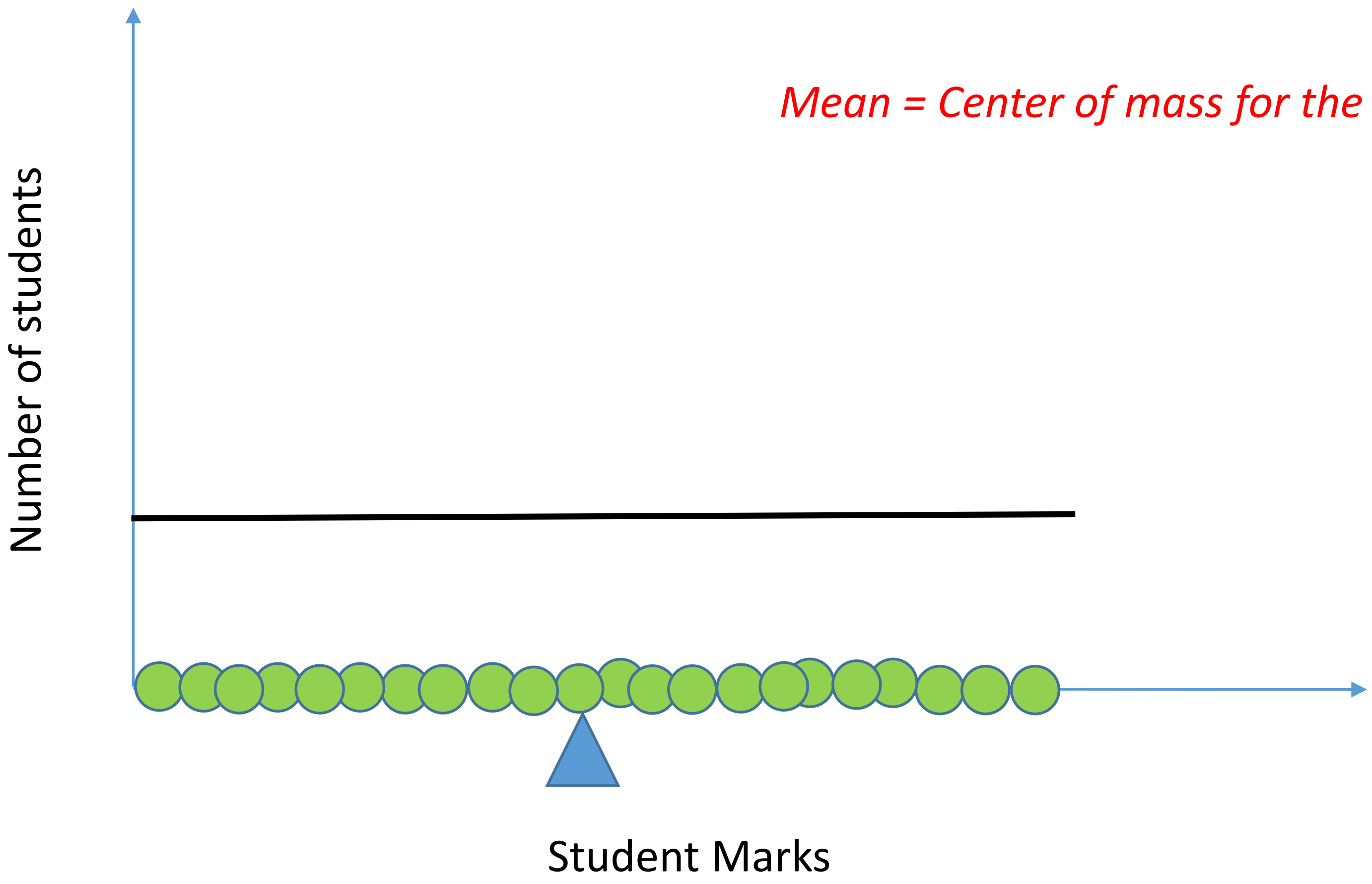
Collecting Probabilities



Two Important Measures of a Random Variable: **Mean and Variance**

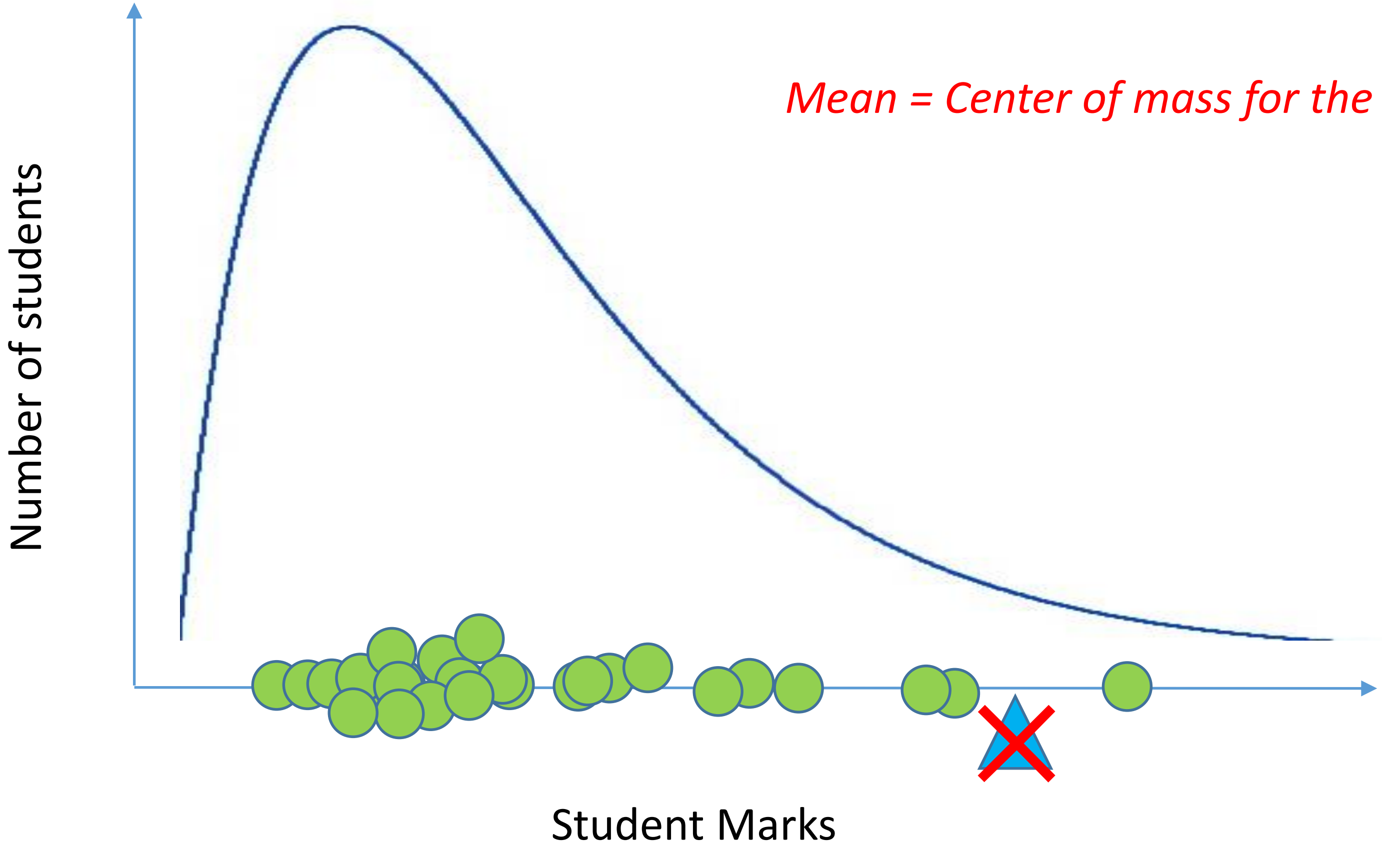
- Next we will talk about two important measures of a random variable that are often used in analysis and design instead of the full probability functions.
- These are *Mean* and *Variance*

Many names: Mean/Average/Expectation



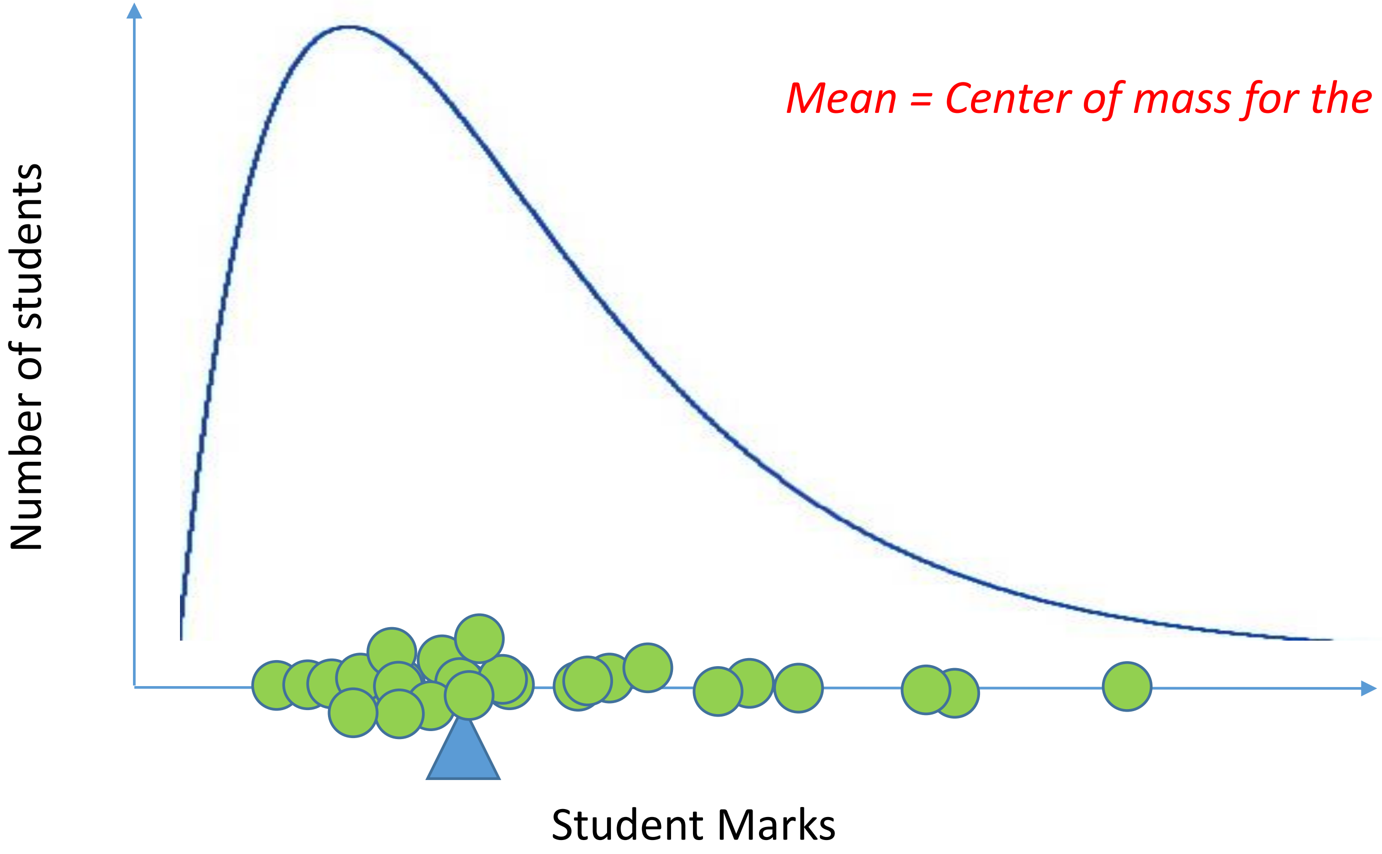
Mean = Center of mass for the distribution (balancing point)

Many names: Mean/Average/Expectation



Mean = Center of mass for the distribution (balancing point)

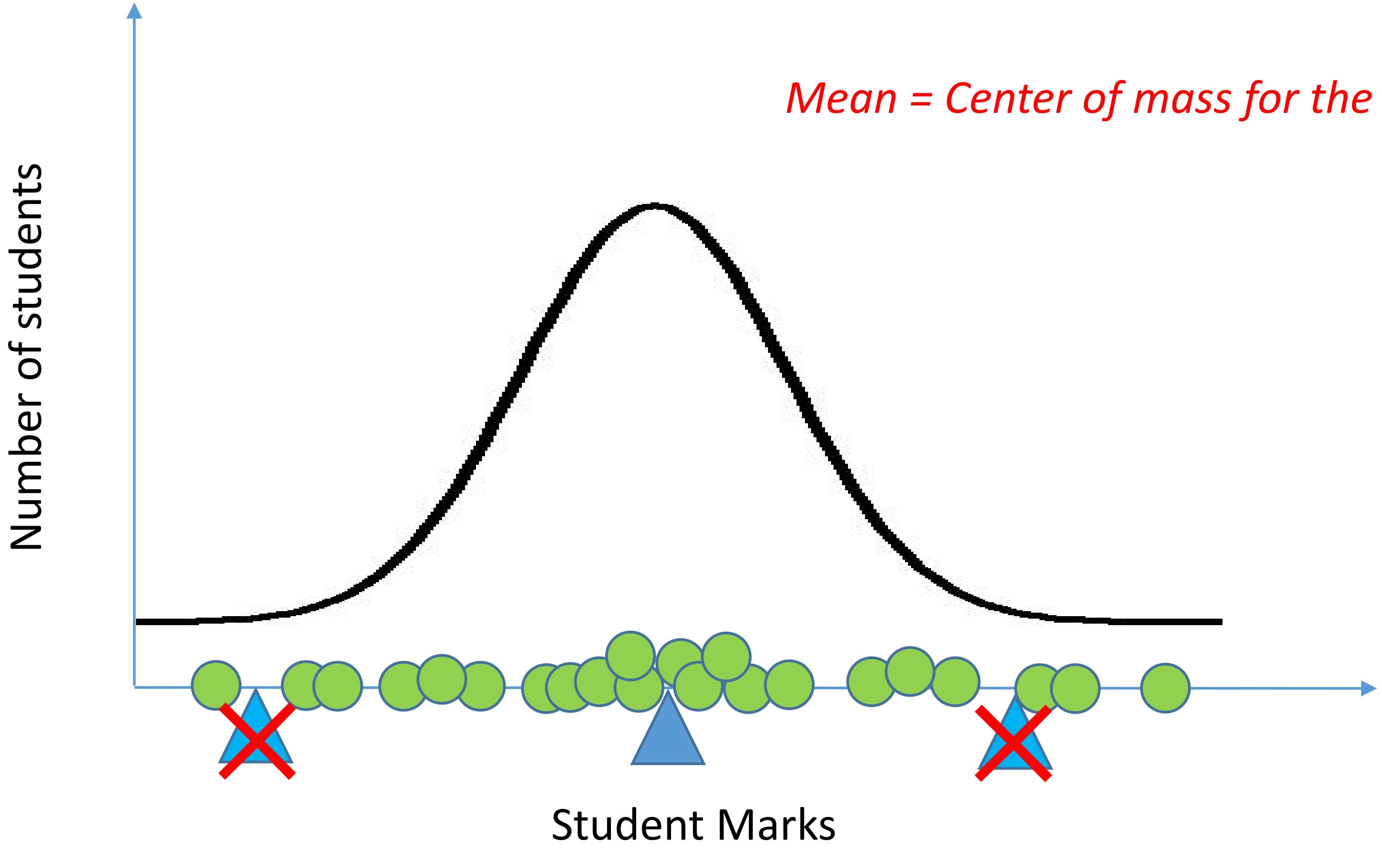
Many names: Mean/Average/Expectation



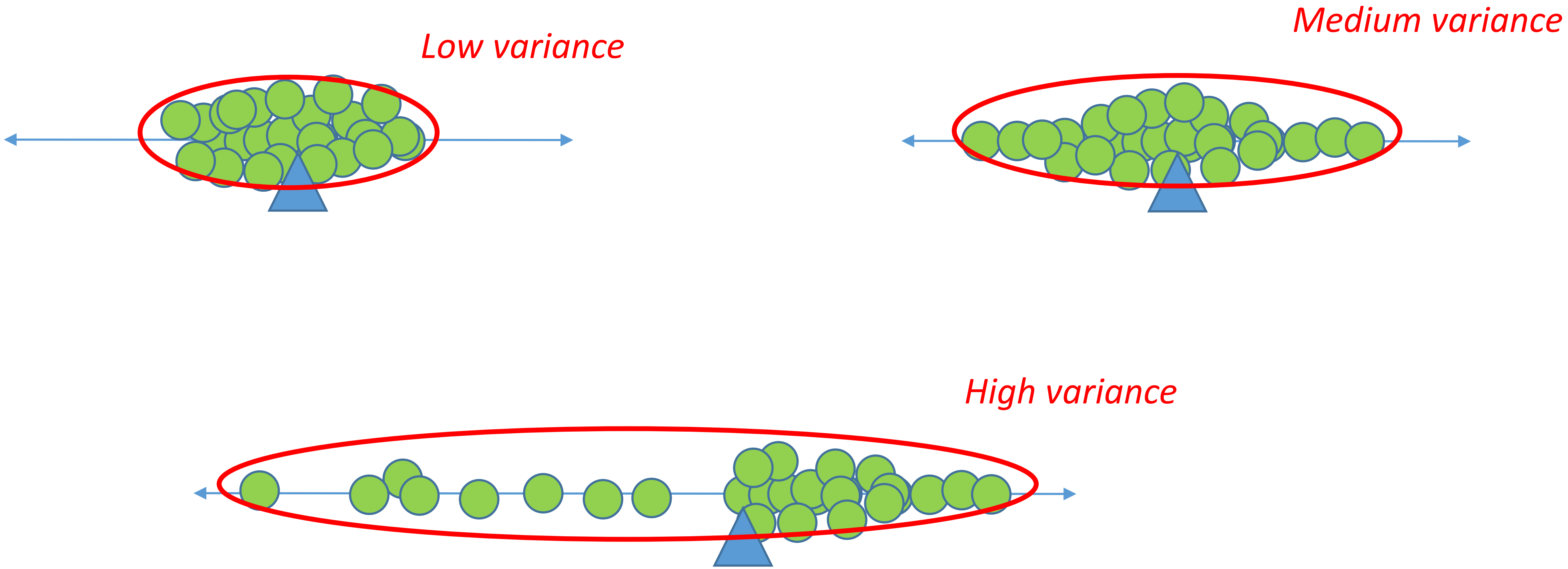
Mean = Center of mass for the distribution (balancing point)

Many names: Mean/Average/Expectation

Mean = Center of mass for the distribution (balancing point)



Variance = degree of spread (how much variation is there in the data/outcomes?)



How shall we write Mean and Variance mathematically?

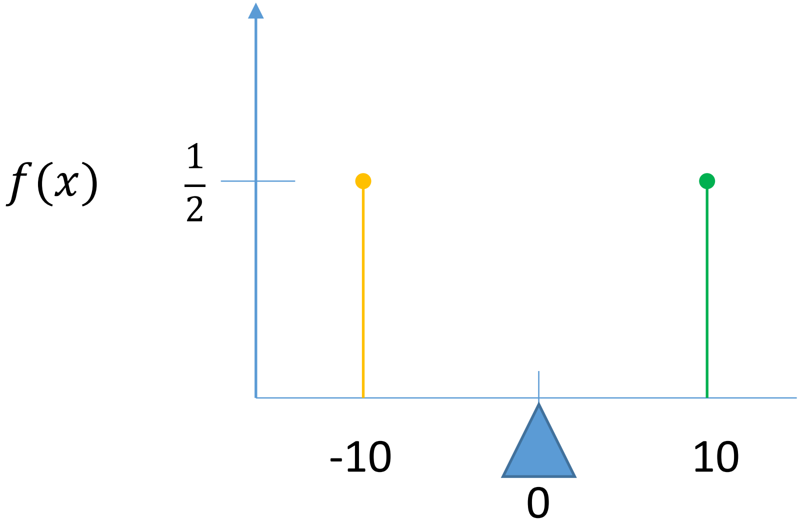
Mean: Let’s play a game …

- Suppose I invite you to a game of toss with following rules
  - We toss several times (e.g., 100 tosses)
  - Every time H appears I give you 10 Riyals
  - Every time T appears I take from you 10 Riyals
- How much can you ***expect*** to win ***on average***?
  - Is there any point in playing this game?

Mean: Let’s play a game …

$$P(H) = P(T) = \frac{1}{2}$$

$$f(10) = f(-10) = \frac{1}{2}$$



Logically

You win 0 Riyals on average

Mathematically

$$E[X] = x_1 f(x_1) + x_2 f(x_2) = 10 \times \frac{1}{2} + (-10) \times \frac{1}{2} = 0$$

Graphically

Mean = where should you place the wedge so that the “scales” are balanced

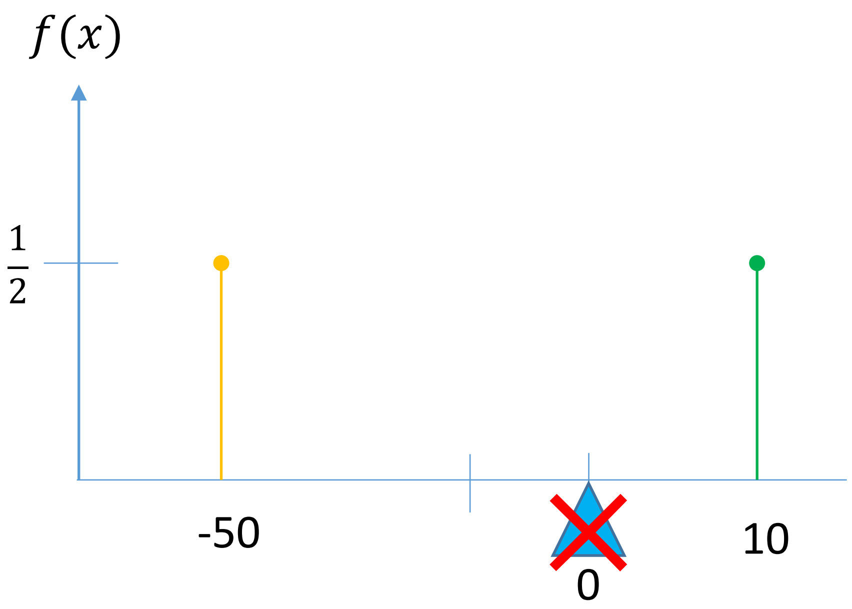
Mean: Let’s play a game …

- Suppose now I change the rules slightly
  - We toss several times (e.g., 100 tosses)
  - Every time H appears I give you 10 Riyals
  - Every time T appears I take from you 50 Riyals
- Now how much can you ***expect*** to win ***on average***?
  - Should you play such a game?

Mean: Let’s play a game …

$$P(H) = P(T) = \frac{1}{2}$$

$$f(10) = f(-50) = \frac{1}{2}$$



On average, you stand to lose 20 Riyals per toss

Logically

You will expect to lose more money

Mathematically

$$E[X] = x_1 f(x_1) + x_2 f(x_2) = 10 \times \frac{1}{2} + (-50) \times \frac{1}{2} = -20$$

Graphically

Mean = where should you place the wedge so that the “scales” are balanced

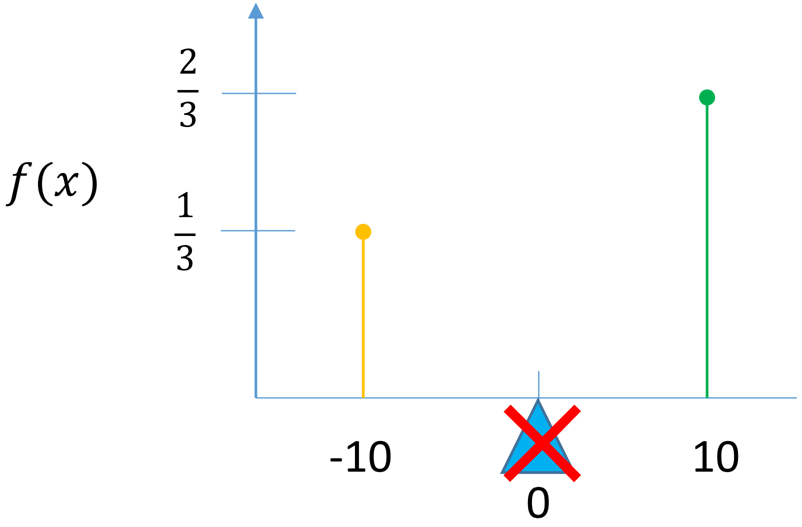
Mean: Let’s play a game …

- Suppose we change the game to rolling a die
  - We roll several times (e.g., 100 rolls)
  - Every time a number above 2 appears I give you 10 Riyals
  - Every time a number less or equal to 2 appears I take from you 10 Riyals
- Now how much can you ***expect*** to win ***on average***?
  - Should you play such a game?

Mean: Let's play a game …

$P(X > 2) = \frac{4}{6} = \frac{2}{3}$ $\Downarrow$ $f(10) = \frac{2}{3}$

$P(X \leq 2) = \frac{2}{6} = \frac{1}{3}$ $\Downarrow$ $f(-10) = \frac{1}{3}$



Logically

You expect to win more

Mathematically

$$E[X] = x_1 f(x_1) + x_2 f(x_2) = 10 \times \frac{2}{3} + (-10) \times \frac{1}{3} = 3.33$$

Graphically

Mean = where should you place the wedge so that the "scales" are balanced

So,

The mean or expected value of a discrete random variable X, denoted μ_X or $E(X)$, is given by

$$\mu_X = E(X) = \sum_x xf(x)$$

- Let’s assume there is a bag with three balls in it with numbers 1, 3, and 5 written on them. You draw one ball at random.
    - X = number written on the ball (random variable)
    - x = 1, 3, 5 (possible values of X)
    - $f(x) = P(X = x)$

x	$f(x)$
1	$\frac{1}{3}$
3	$\frac{1}{3}$
5	$\frac{1}{3}$

$$\mu_X = E(X) = \sum_x x f(x) = \left(1 \times \frac{1}{3}\right) + \left(3 \times \frac{1}{3}\right) + \left(5 \times \frac{1}{3}\right) = 3$$

- Let’s assume there is a bag with three balls in it with numbers 1, 3, and 5 written on them. You draw one ball at random.
  - X = number written on the ball (random variable)
  - x = 1, 3, 5 (possible values of X)
  - $f(x) = P(X = x)$

x		$f(x)$
1	×	$\frac{1}{3}$
3	×	$\frac{1}{3}$
5	×	$\frac{1}{3}$

$$\mu_X = E(X) = \sum_x x f(x) = \left(1 \times \frac{1}{3}\right) + \left(3 \times \frac{1}{3}\right) + \left(5 \times \frac{1}{3}\right) = 3$$

+ +

Variance

The variance of a discrete random variable X, denoted σ_X^2 or $V(X)$, is given by

$$\sigma_X^2 = V(X) = E(X - \mu_X)^2$$

Several equivalent formulations of variance

$$\boxed{\sigma_X^2 = E(X - \mu_X)^2}$$

$$\boxed{\sigma_X^2 = \sum_x (x - \mu_X)^2 f(x)}$$

$$\boxed{\sigma_X^2 = E(X^2) - \mu_X^2}$$

$$\boxed{\sigma_X^2 = \sum_x x^2 f(x) - \mu_X^2}$$

- Let’s assume there is a bag with three balls in it with numbers 1, 3, and 5 written on them. You draw one ball at random.
  - X = number written on the ball (random variable)
  - x = 1, 3, 5 (possible values of X)
  - $f(x) = P(X = x)$

x	$f(x)$
1	$\frac{1}{3}$
3	$\frac{1}{3}$
5	$\frac{1}{3}$

$$\mu_X = E(X) = \sum_x x f(x) = \left(1 \times \frac{1}{3}\right) + \left(3 \times \frac{1}{3}\right) + \left(5 \times \frac{1}{3}\right) = 3$$

$$\sigma_X^2 = \sum_x x^2 f(x) - \mu_X^2 = \left((1)^2 \times \frac{1}{3}\right) + \left((3)^2 \times \frac{1}{3}\right) + \left((5)^2 \times \frac{1}{3}\right) - (3)^2 = 2.66$$

Standard Deviation

The square root of variance is called standard deviation, and denoted by σ_X

$$\sigma_X = \sqrt{V(X)}$$

- Let’s assume there is a bag with three balls in it with numbers 1, 3, and 5 written on them. You draw one ball at random.
  - X = number written on the ball (random variable)
  - x = 1, 3, 5 (possible values of X)
  - $f(x) = P(X = x)$

$\boldsymbol{x}$	$\boldsymbol{f(x)}$
1	$\frac{1}{3}$
3	$\frac{1}{3}$
5	$\frac{1}{3}$

$$\mu_X = E(X) = \sum_x xf(x) = \left(1 \times \frac{1}{3}\right) + \left(3 \times \frac{1}{3}\right) + \left(5 \times \frac{1}{3}\right) = 3$$

$$\sigma_X^2 = \sum_x x^2 f(x) - \mu_X^2 = \left((1)^2 \times \frac{1}{3}\right) + \left((3)^2 \times \frac{1}{3}\right) + \left((5)^2 \times \frac{1}{3}\right) - (3)^2 = 2.66$$

$$\sigma_X = \sqrt{2.66} = 1.63$$

What is the mean of a constant (or a non-random variable)?

What is the mean of a constant (or a non-random variable)?

If C is a constant, then

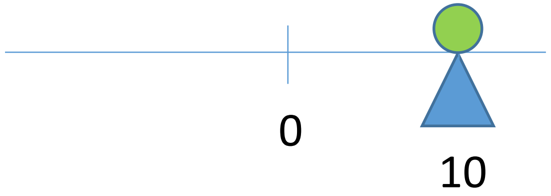
$$\boxed{E(C) = C}$$

e.g. $E(10) = 10$ ← A constant takes only one value, so the "balancing point" is that value itself

If Z_t is a non-random variable, then

$$\boxed{E(Z_t) = Z_t}$$

e.g. if $Z_t = t$, then we already know that $Z_1 = 1$, $Z_2 = 2$, ... (i.e., the values of Z_t are non-random)



What is the variance of a constant (or a non-random variable)?

What is the variance of a constant (or a non-random variable)?

If C is a constant, then

$$\boxed{V(C) = 0}$$

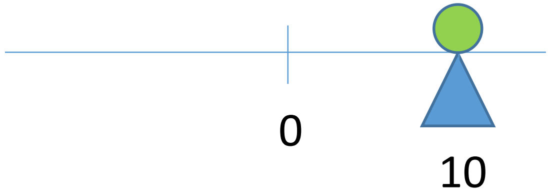
e.g. $V(10) = 0$

A constant takes only one value, so there is no “spread”

If Z_t is a non-random variable, then

$$\boxed{V(Z_t) = 0}$$

e.g. if $Z_t = t$, then we already know that $Z_1 = 1$, $Z_2 = 2$, ... (i.e., the values of Z_t are non-random), and so we know that $V(Z_1) = V(1) = 0, V(Z_2) = V(2) = 0$, ...



Some important properties of the mean

If *X* is a random variable and ***a* and *b* are constants**, then

$$\boxed{E(aX+b)=aE(X)+b}$$

For **any two random variables** *X* and *Y*, we have

$$\boxed{\begin{array}{l} E(X+Y)=E(X)+E(Y) \\ E(X-Y)=E(X)-E(Y) \end{array}}$$

For **two independent random variables** *X* and *Y*, we have

$$\boxed{E(XY)=E(X)E(Y)}$$

Some important properties of the mean

If *X* is a random variable and ***v(X)*** is some function of X, then

$$\boxed{E(v(X)) = \sum_{x} v(x) f(x)}$$

For example, if $v(X) = X^3$, then

$$\boxed{E(v(X)) = E(X^3) = \sum_{x} x^3 f(x)}$$

Some important properties of variance

If ***X and Y*** **are any two random variables** and ***a, b*** **are constants**, then

$$\boxed{V(aX + bY) = a^2V(X) + b^2V(Y) + 2abE((X - \mu_X)(Y - \mu_Y))}$$

If ***X and Y*** **are two independent random variables** and ***a, b are*** **constants**, then

$$\boxed{\begin{aligned} V(aX + bY) &= a^2V(X) + b^2V(Y) \\ V(aX - bY) &= a^2V(X) + b^2V(Y) \end{aligned}}$$

Example 1

Consider two zero-mean unit variance random variables X, Y. Define a new random variable W = 2X+Y. Given that X and Y are independent, find mean and variance of W.

We are given that

$$\mu_X = \mu_Y = 0$$
$$\sigma_X^2 = \sigma_Y^2 = 1$$

And

$$X, Y \text{ are independent}$$

Therefore

$$\mu_W = E(W) = E(2X + Y) = 2E(X) + E(Y) = 0$$
$$V(W) = V(2X + Y) = 4\mathrm{V}(\mathrm{X}) + \mathrm{V}(\mathrm{Y}) = 4 + 1 = 5$$

Example 2

Consider two zero-mean unit variance random variables X, Z. Define a new random variable W = 2X+Z. Given that E(XZ) = 1, find mean and variance of W.

We are given that

$$\mu_X = \mu_Z = 0$$
$$\sigma_X^2 = \sigma_Z^2 = 1$$

Therefore

$$\mu_W = E(W) = E(2X + Z) = 2E(X) + E(Z) = 0$$
$$V(W) = V(2X + Z) = 4\mathrm{V}(\mathrm{X}) + \mathrm{V}(\mathrm{Z}) + 2(2)\mathrm{E}\big((\mathrm{X} - \mu_X)(\mathrm{Z} - \mu_{\mathrm{Z}})\big) = 4 + 1 + 4(1 + 0 + 0 + 0) = 9$$

Questions?? Thoughts??

EE 302
Probabilistic Methods in Electrical Engineering

with

Dr. Naveed R. Butt

@

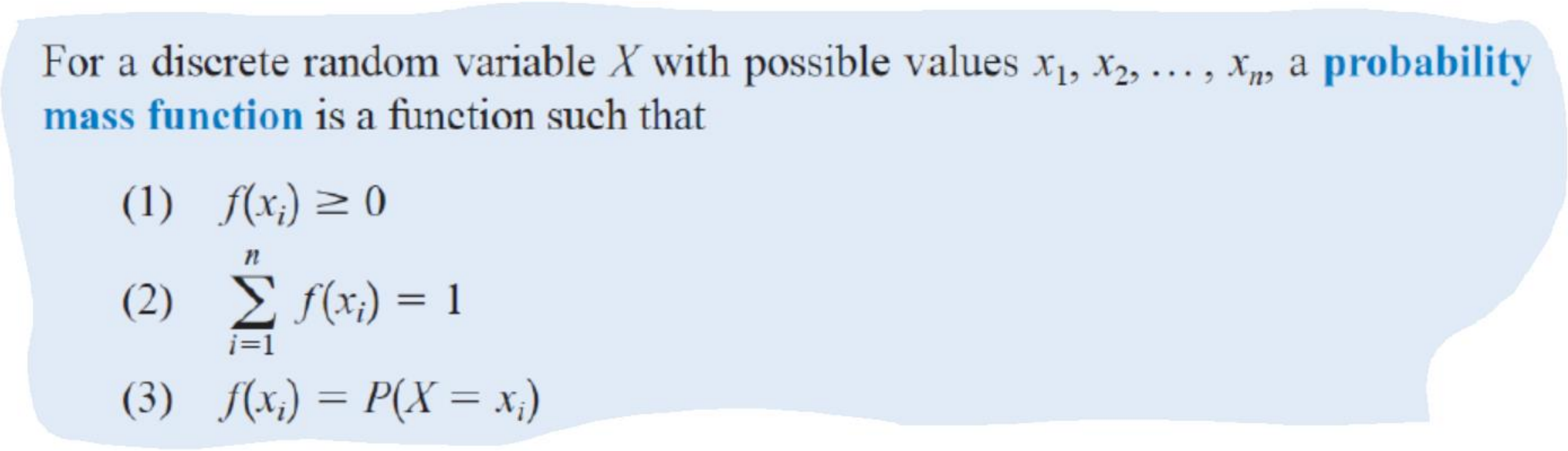
Jouf University

We have previously talked about …

Probability Mass Function (PMF): $f(x)$

$$f(x) = P(X = x)$$

Probability that the discrete random variable X takes the value x



For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a **probability mass function** is a function such that

(1) $f(x_i) \geq 0$

(2) $\sum_{i=1}^{n} f(x_i) = 1$

(3) $f(x_i) = P(X = x_i)$

What are some of the common distribution functions for discrete random variables?

- Next, we will look at some probability distributions ($f(x)$), that are often used in practice for discrete random variables
- These include
  - Uniform Distribution
  - Binomial Distribution
  - Geometric Distribution
  - Negative Binomial Distribution
  - Poisson Distribution

Based on **Bernoulli Trials**

Discrete Uniform Distribution = when all possible outcomes have equal probabilities

- X = result of a rolling a die
- $x = 1, 2, 3, 4, 5, 6$ (possible values of X)
- All outcomes have equal probability
- $f(x) = \begin{cases} \frac{1}{6}, & x = 1,2,3,4,5,6 \\ 0\,, & otherwise \end{cases}$

Discrete Uniform Distribution = when all possible outcomes have equal probabilities

A random variable X has a discrete uniform distribution if it has a finite set of outcomes (say $x_1, x_2, x_3, \ldots, x_n$), and each outcome has exactly the same probability. The distribution of X with n equally probably outcomes may then be written as

$$f(x) = \begin{cases} \frac{1}{n}, & x = x_1, x_2, x_3, \ldots, x_n \\ 0\,, & otherwise \end{cases}$$

Discrete Uniform Distribution = when all possible outcomes have equal probabilities

Mean and Variance of a uniformly distributed random variable X

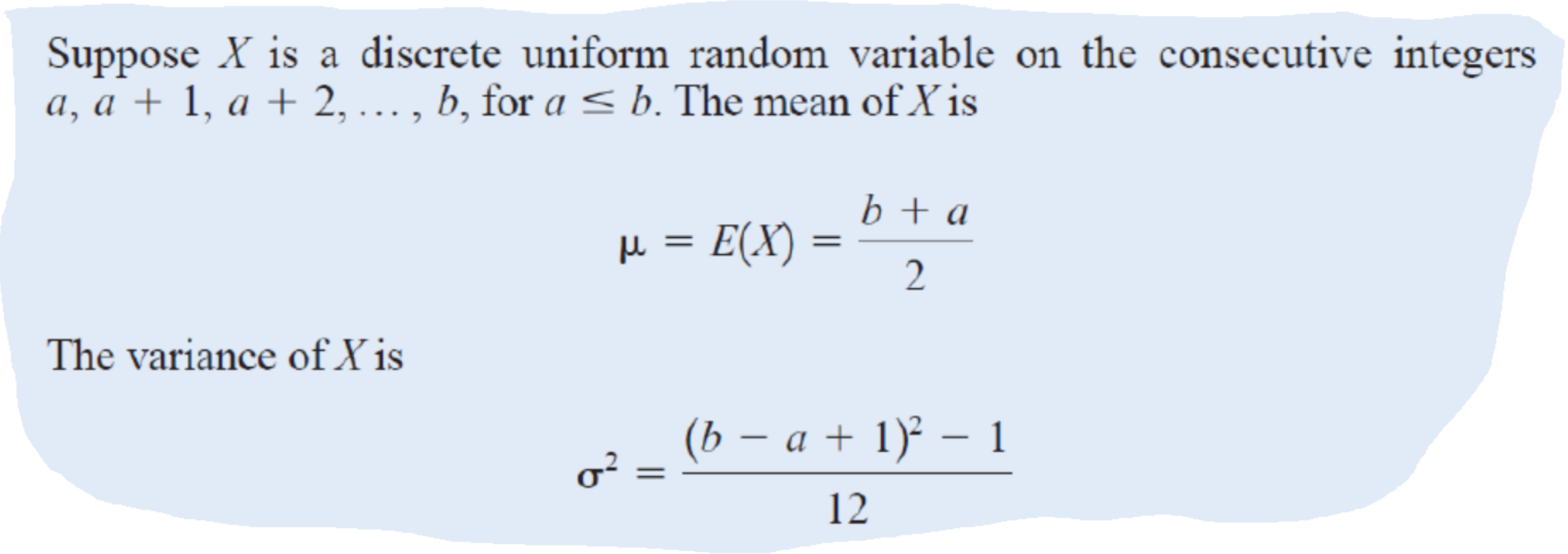
$$f(x) = \begin{cases} \frac{1}{n}, & x = x_1, x_2, x_3, \dots, x_n \\ 0\,, & otherwise \end{cases}$$

$$E[X] = \mu_X = \sum_x x f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$V[X] = \sum_x (x - \mu_X)^2 f(x) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu_X)^2$$

Discrete Uniform Distribution

Special case: when x_i are integers in range $[a\ b]$



Suppose X is a discrete uniform random variable on the consecutive integers $a, a + 1, a + 2, \ldots, b$, for $a \leq b$. The mean of X is

$$\mu = E(X) = \frac{b + a}{2}$$

The variance of X is

$$\sigma^2 = \frac{(b - a + 1)^2 - 1}{12}$$

Example

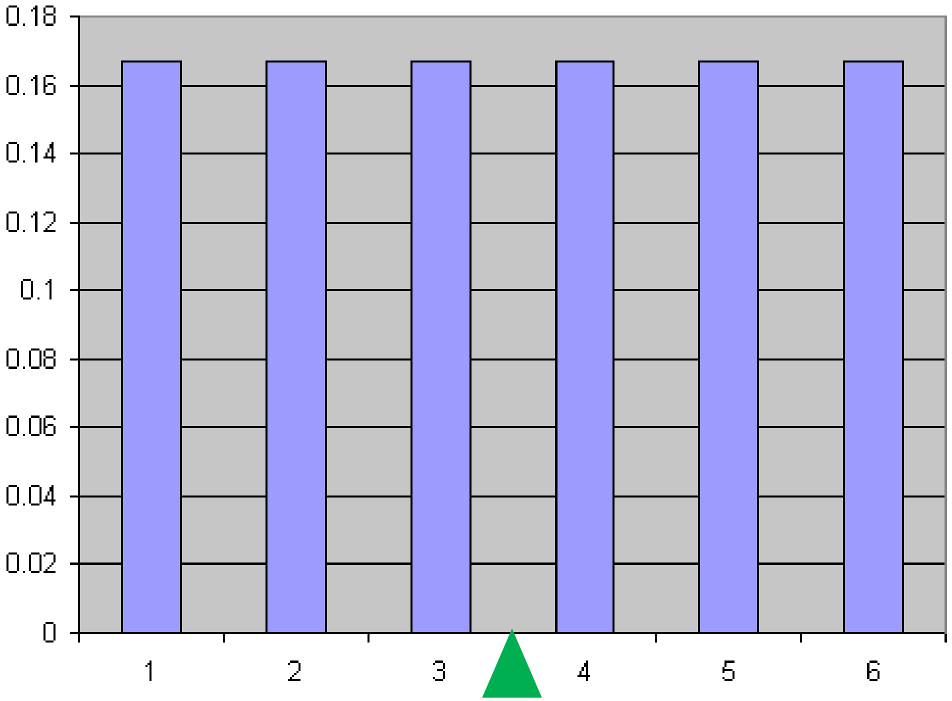
If $$f(x) = \begin{cases} \frac{1}{6}, & x = 1, 2, 3, 4, 5, 6 \\ 0\,, & otherwise \end{cases}$$

then

$$a = 1, \qquad b = 6$$

$$E[X] = \frac{6+1}{2} = 3.5$$

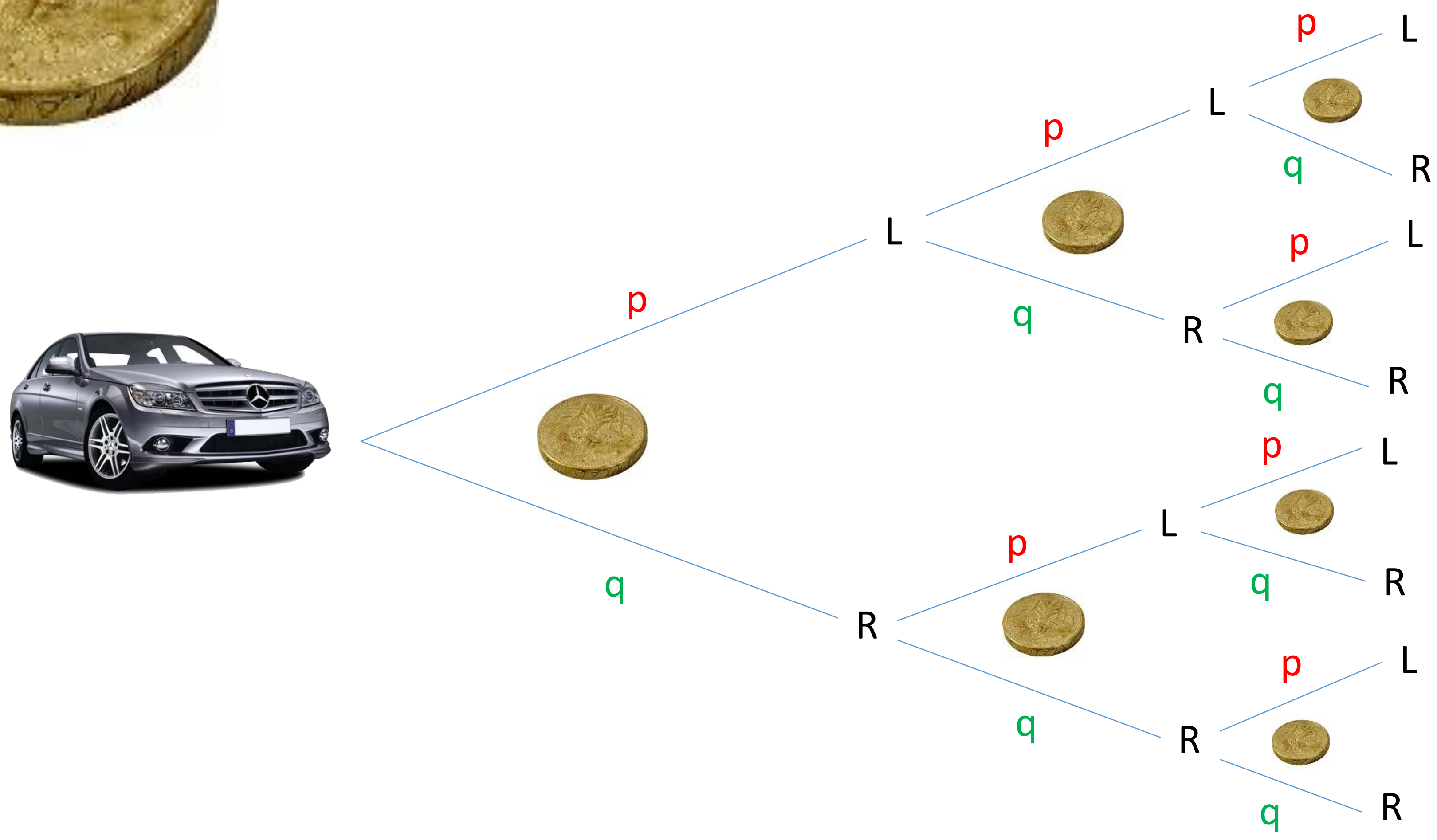
$$V[X] = \frac{(6-1+1)^2 - 1}{12} = \frac{35}{12} = 2.92$$



What are some of the common distribution functions for discrete random variables?

- Next, we will look at some probability distributions ($f(x)$), that are often used in practice for discrete random variables
- These include
    - Uniform Distribution
    - Binomial Distribution
    - Geometric Distribution
    - Negative Binomial Distribution
    - Poisson Distribution

Based on **Bernoulli Trials**



LLL

LLR

LRL

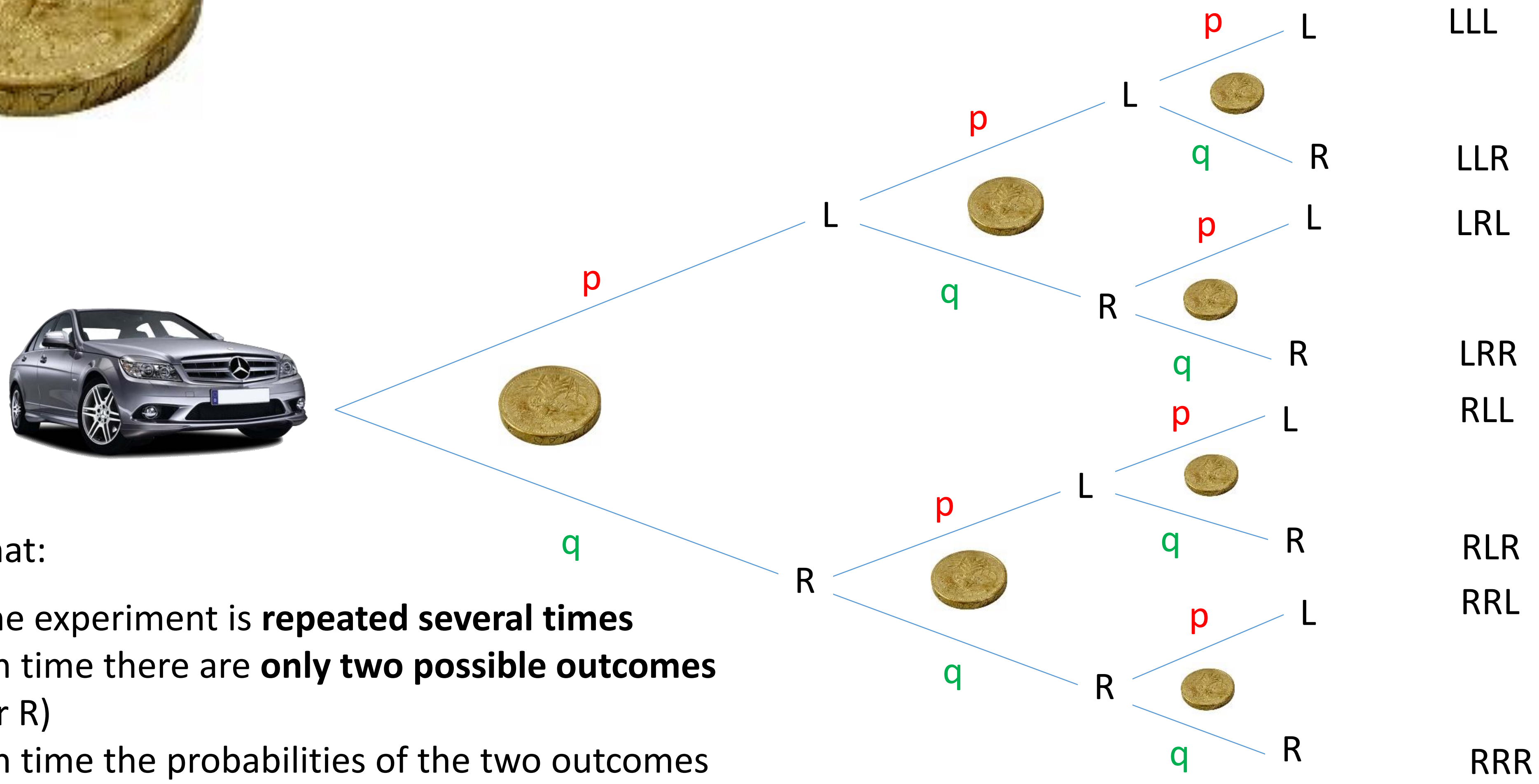
LRR

RLL

RLR

RRL

RRR



Note that:

- Same experiment is **repeated several times**
- Each time there are **only two possible outcomes** (L or R)
- Each time the probabilities of the two outcomes are the same (***P(L) = p, P(R)= q***)
- All the **repetitions are independent** of each other

Bernoulli Process & Bernoulli Trials

- An experiment that satisfies the conditions we saw in the previous example is called a **Bernoulli Process**, and each repetition is called a **Bernoulli Trial**.

In summary, a Bernoulli Process is a random process
that satisfies the following conditions

1. It consists of **several repetitions**
2. Each repetition is **independent** of the other
3. Each repetition has **only two possible outcomes** (*usually called "**success**" and "**failure**"*)
4. The **probability of "success" is the same** during each repetition (*common notation* $\boldsymbol{p = P(\text{"success"}), q = P("failure")}$)

Some Questions we could ask

- What is the probability that in the **n repetitions** you take **exactly x left turns**?
- What is the probability that the **first left turn** you take is **at repetition number x**?
- What is the probability that it takes you **x repetitions to take exactly r left turns**?

Each of these leads to a different random variable x, with its own probability mass function $f(x)$

What is the probability that you take three left turns?

Probability (first left **and** second left **and** third left) = ?

Let t_1 = first turn, t_2 = second turn, t_3 = third turn

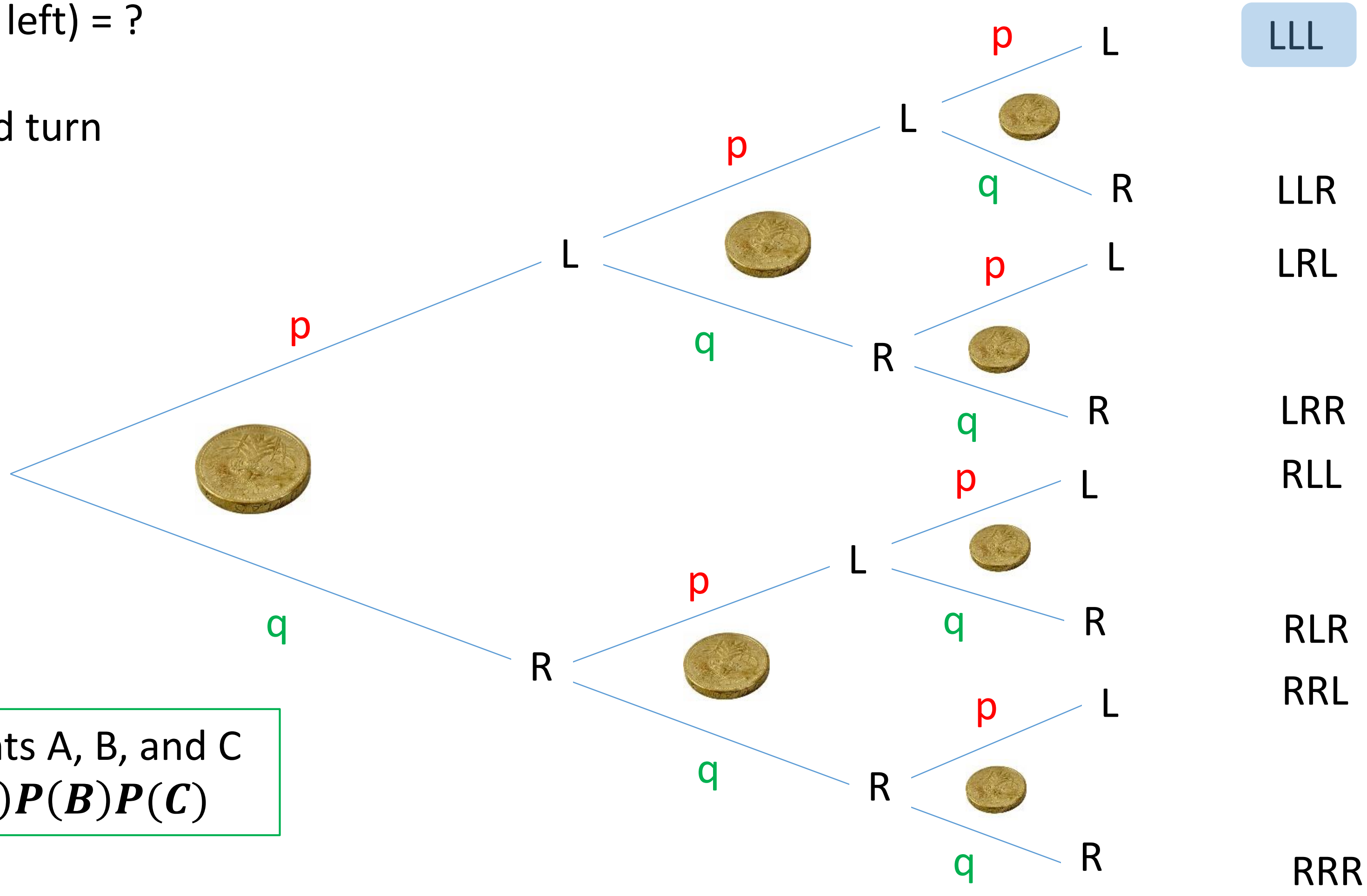
Then,

$$P(t_1 = L \cap t_2 = L \cap t_3 = L) = ?$$

$$= P(t_1 = L)P(t_2 = L)P(t_3 = L)$$

$$= p \times p \times p = p^3$$

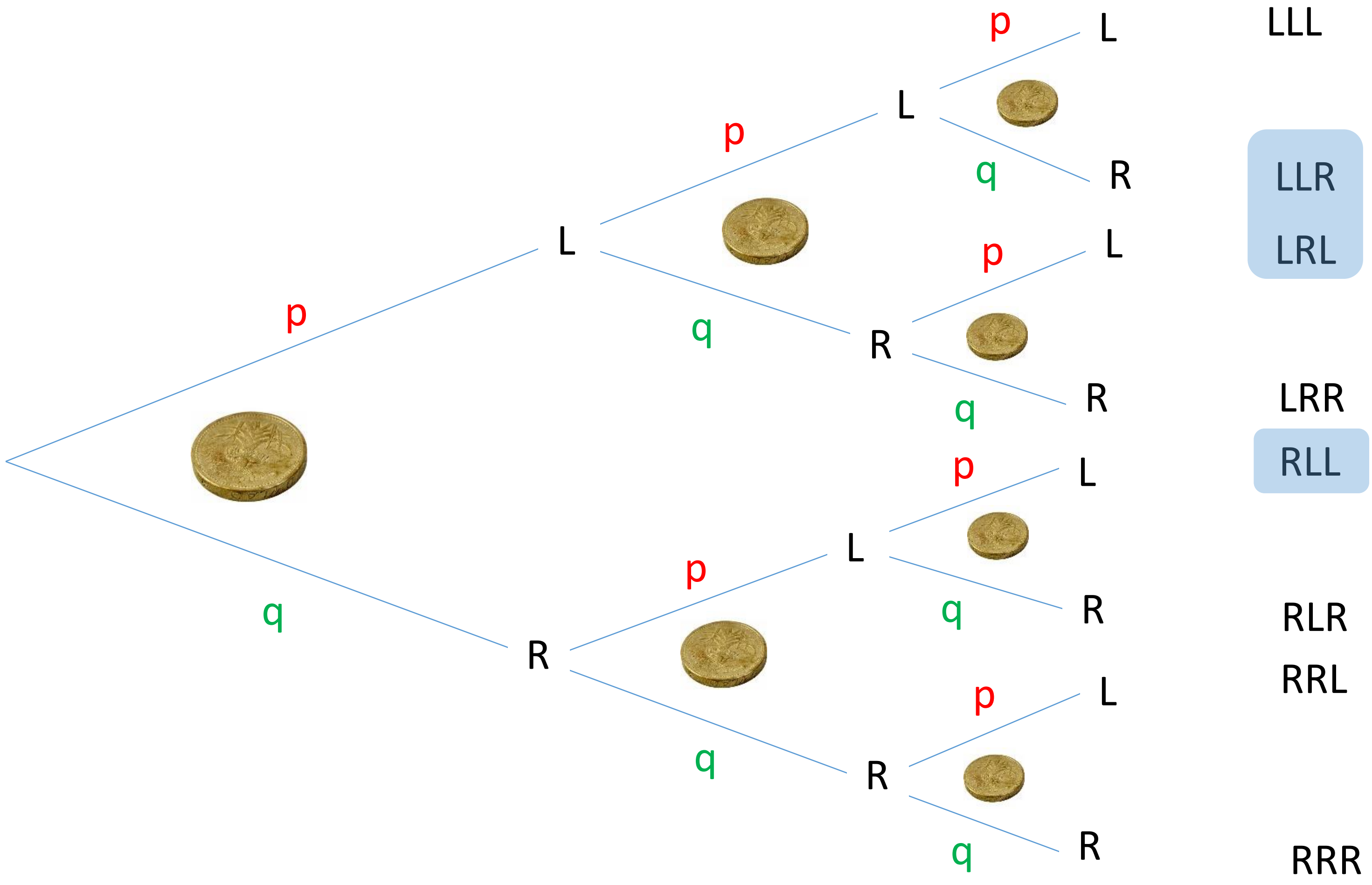
Recall that for **independent** events A, B, and C we have $\boldsymbol{P(A \cap B \cap C) = P(A)P(B)P(C)}$



What is the probability that you take exactly two left turns?

$$P(t_1 = L \cap t_2 = L \cap t_3 = R)$$
$$+P(t_1 = L \cap t_2 = R \cap t_3 = L)$$
$$+P(t_1 = R \cap t_2 = L \cap t_3 = L)$$

$$= ppq + pqp + qpp \;= 3p^2q$$



Next, We will use this understanding to answer several such questions, and develop various probability distributions

What are some of the common distribution functions for discrete random variables?

- Next, we will look at some probability distributions ($f(x)$), that are often used in practice for discrete random variables
- These include
  - Uniform Distribution
  - Binomial Distribution
  - Geometric Distribution
  - Negative Binomial Distribution
  - Poisson Distribution

Based on **Bernoulli Trials**

Binomial Distribution

Consider an experiment consisting of n Bernoulli trials (i.e., independent trials with constant probability of “success” denoted p)

Defining the setup (experiment)

Let $X = \boldsymbol{number\ of\ successes}\ in\ \mathbf{n}\ Bernoulli\ trials$

Defining the random variable

Then the probability distribution of X is given by

$$f(x) = \binom{n}{x} p^x (1 - p)^{n-x} \qquad x = 0, 1, \ldots, n$$

Finding formula for distribution of x

And we say that X has a **Binomial Distribution**

Naming the distribution

Geometric Distribution

Consider an experiment consisting of Bernoulli trials (i.e., independent trials with constant probability of “success” denoted p)

Defining the setup (experiment)

Let $X = \boldsymbol{number\ of\ trials}\ till\ first\ success$

Defining the random variable

Then the probability distribution of X is given by

$$f(x) = (1 - p)^{x-1}p \qquad x = 1, 2, \ldots$$

Finding formula for distribution of x

And we say that X has a **Geometric Distribution**

Naming the distribution

Negative Binomial Distribution

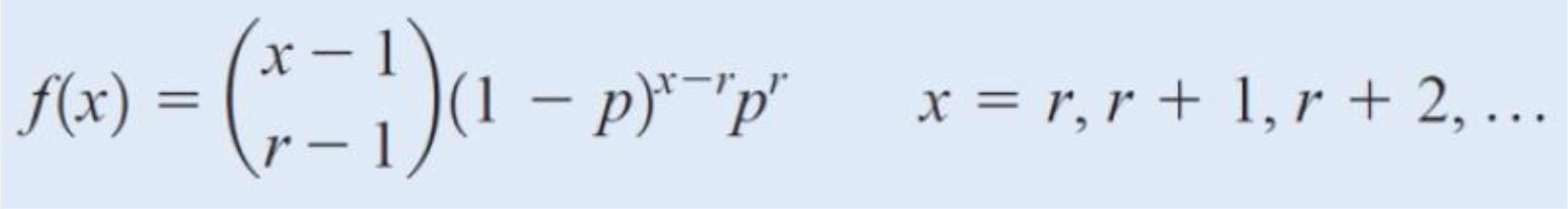
Consider an experiment consisting of Bernoulli trials (i.e., independent trials with constant probability of "success" denoted p)

Defining the setup (experiment)

Let $X = \boldsymbol{number\ of\ trials}\ till\ \boldsymbol{r}\ successes$

Defining the random variable

Then the probability distribution of X is given by



$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^{r} \qquad x = r, r+1, r+2, \ldots$$

Finding formula for distribution of x

And we say that X has a **Negative Binomial Distribution**

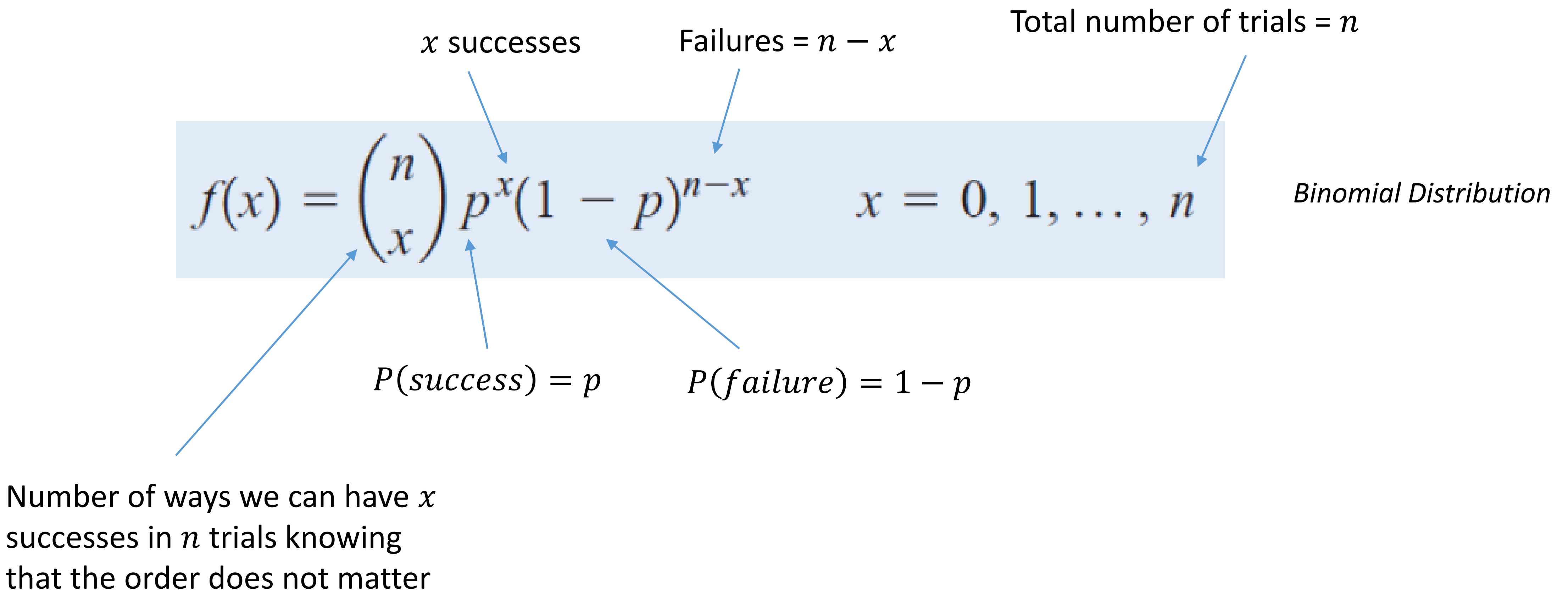
Naming the distribution

Recall …

$$^{n}C_{r} \text{ or } \binom{n}{r} = \frac{n!}{(n-r)!r!}$$

Recall **5! = 5 x 4 x 3 x 2 x 1**
And **0! = 1**

Interpretations …



x successes

Failures = $n - x$

Total number of trials = n

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0, 1, \ldots, n$$

Binomial Distribution

$P(success) = p$

$P(failure) = 1 - p$

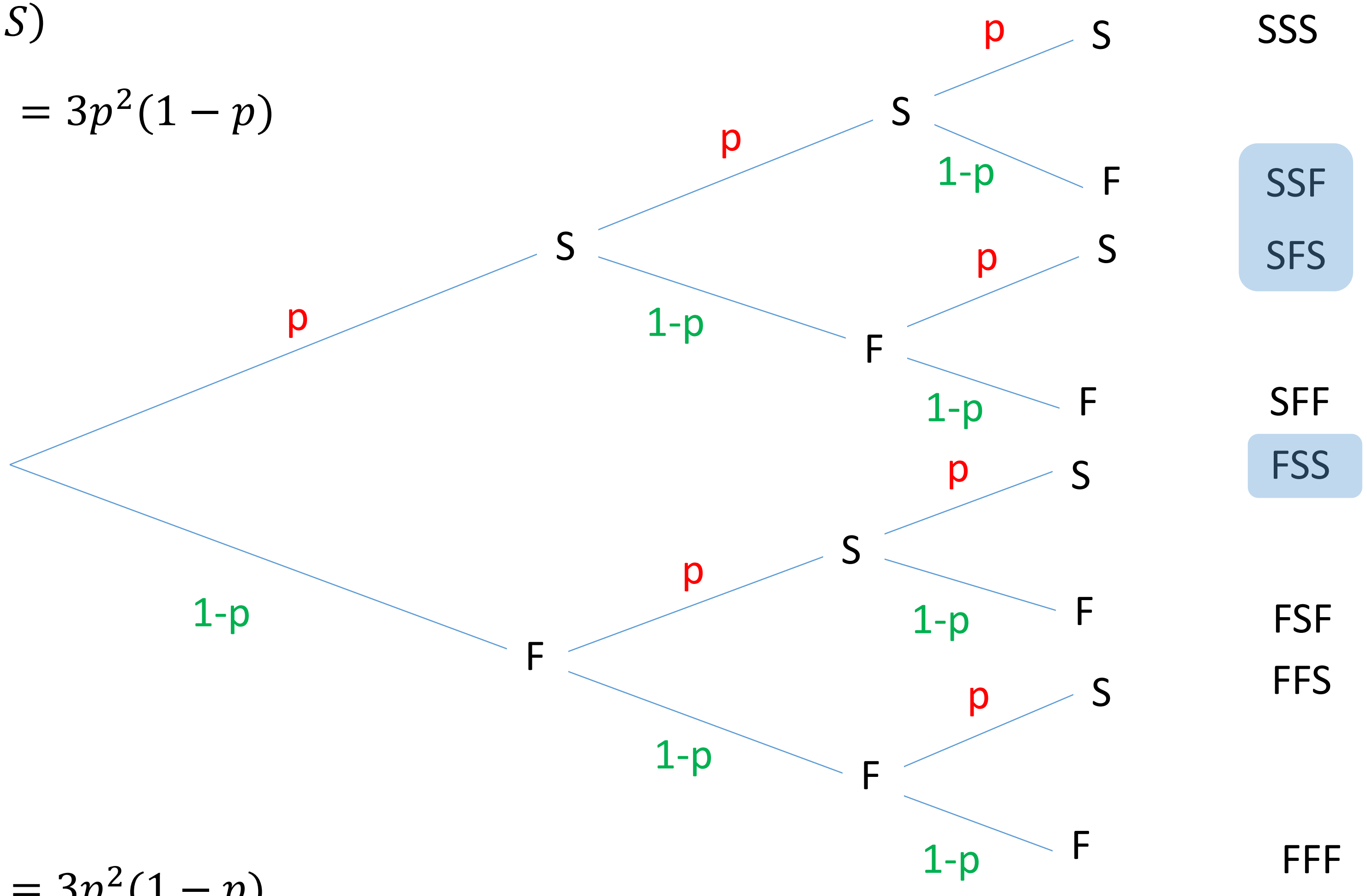
Number of ways we can have x successes in n trials knowing that the order does not matter

$X = \boldsymbol{number\ of\ successes}\ in\ \boldsymbol{n}\ trials$

What is the probability of two success in three trials?

$$P(S \cap S \cap F) + P(S \cap F \cap S) + P(F \cap S \cap S)$$

$$= pp(1-p) + p(1-p)p + (1-p)pp \ = 3p^2(1-p)$$



Using the Binomial Distribution formula with $n = 3, x = 2$ we get the same result

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$f(2) = \binom{3}{2} p^2(1-p) = \frac{3!}{(3-2)!\,2!} p^2(1-p) = 3p^2(1-p)$$

Interpretations …

$P(failure) = 1 - p$

$P(success) = p$

$$f(x) = (1 - p)^{x-1}p \qquad x = 1, 2, \ldots$$

Geometric Distribution

There has to be at least one trial for one success to occur (so no $x = 0$)

There are a total of $x - 1$ failures *before* the first success occurs

Success occurs in trial number x but not before

$X = \boldsymbol{number\ of\ trials}\ till\ first\ success$

What is the probability of that first success occurs in third trial?

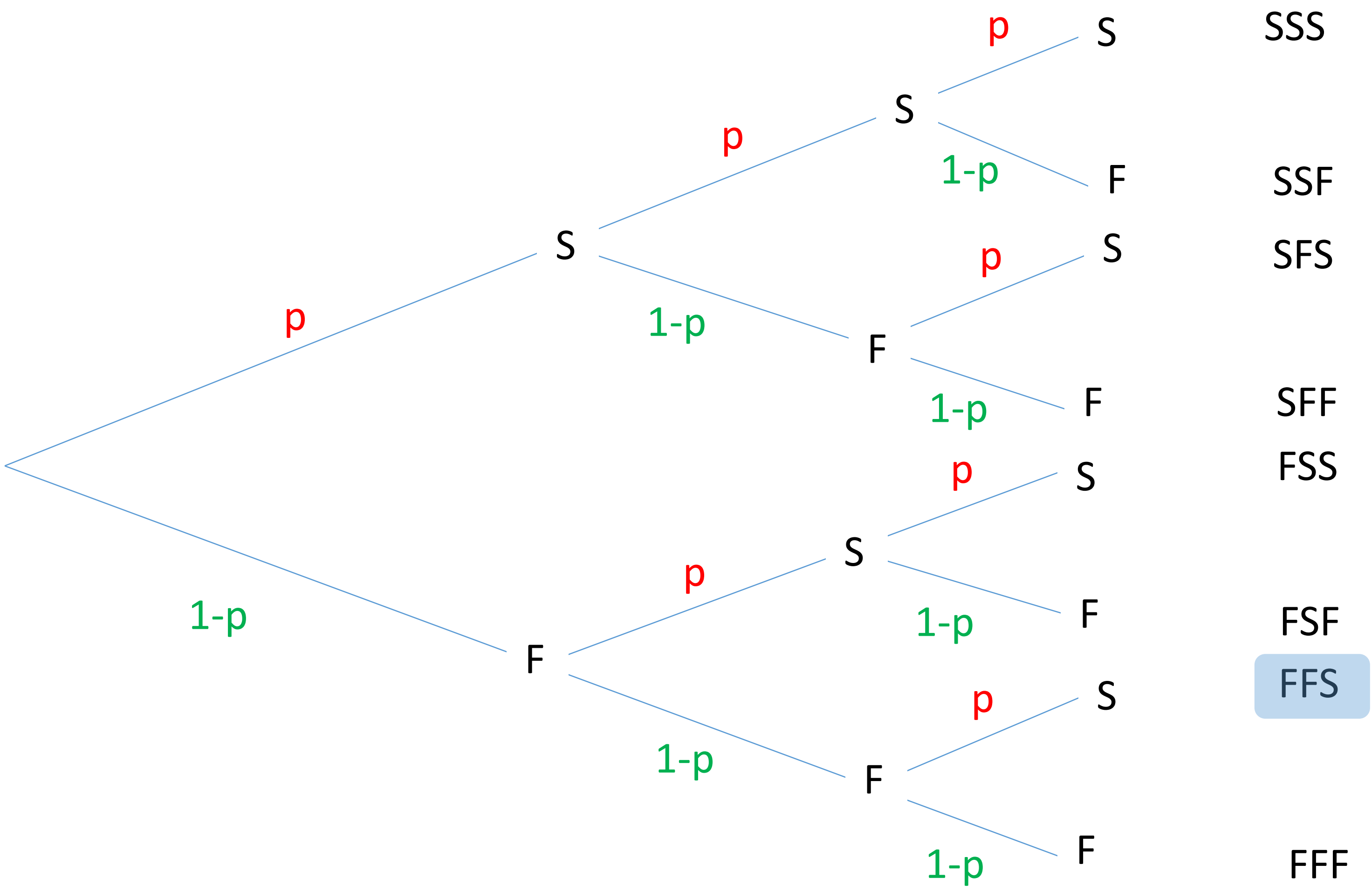
$$P(F \cap F \cap S)$$

$$= (1-p)(1-p)p = (1-p)^2p$$

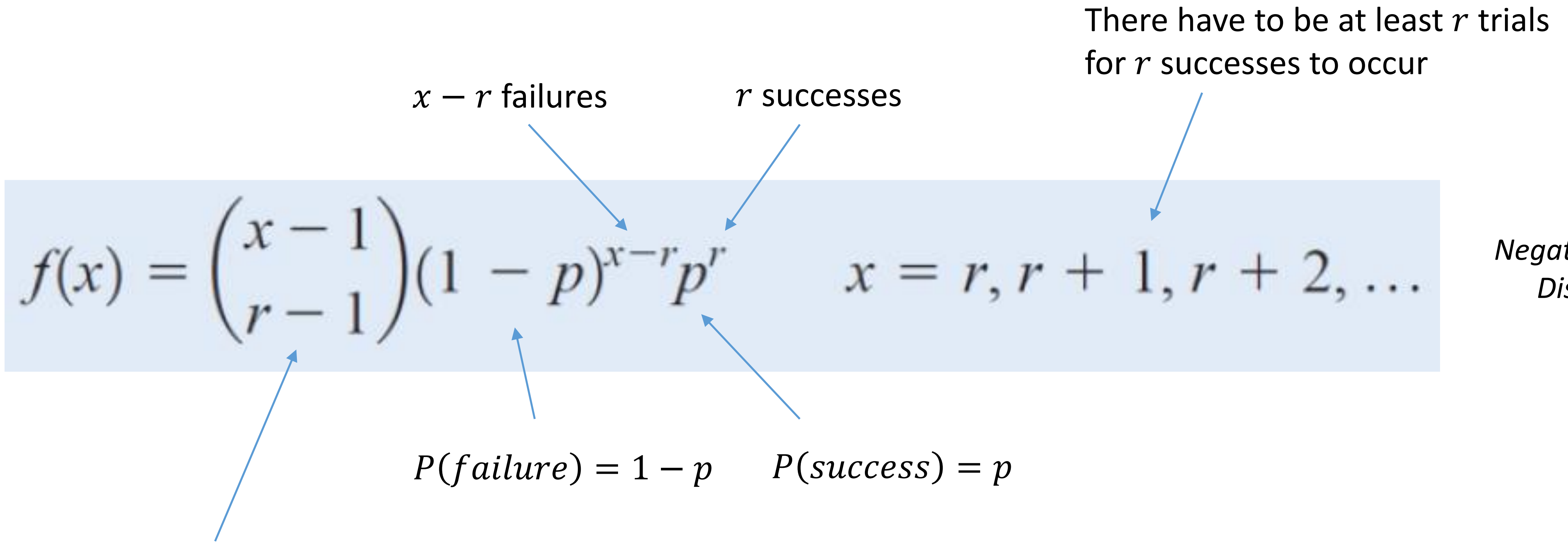
Using the Geometric Distribution formula with $x = 3$ we get the same result

$$f(x) = (1-p)^{x-1}p$$

$$f(3) = (1-p)^2p$$



Interpretations …



There have to be at least r trials for r successes to occur

$x - r$ failures

r successes

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r \qquad x = r, r+1, r+2, \ldots$$

Negative Binomial Distribution

$P(failure) = 1 - p$

$P(success) = p$

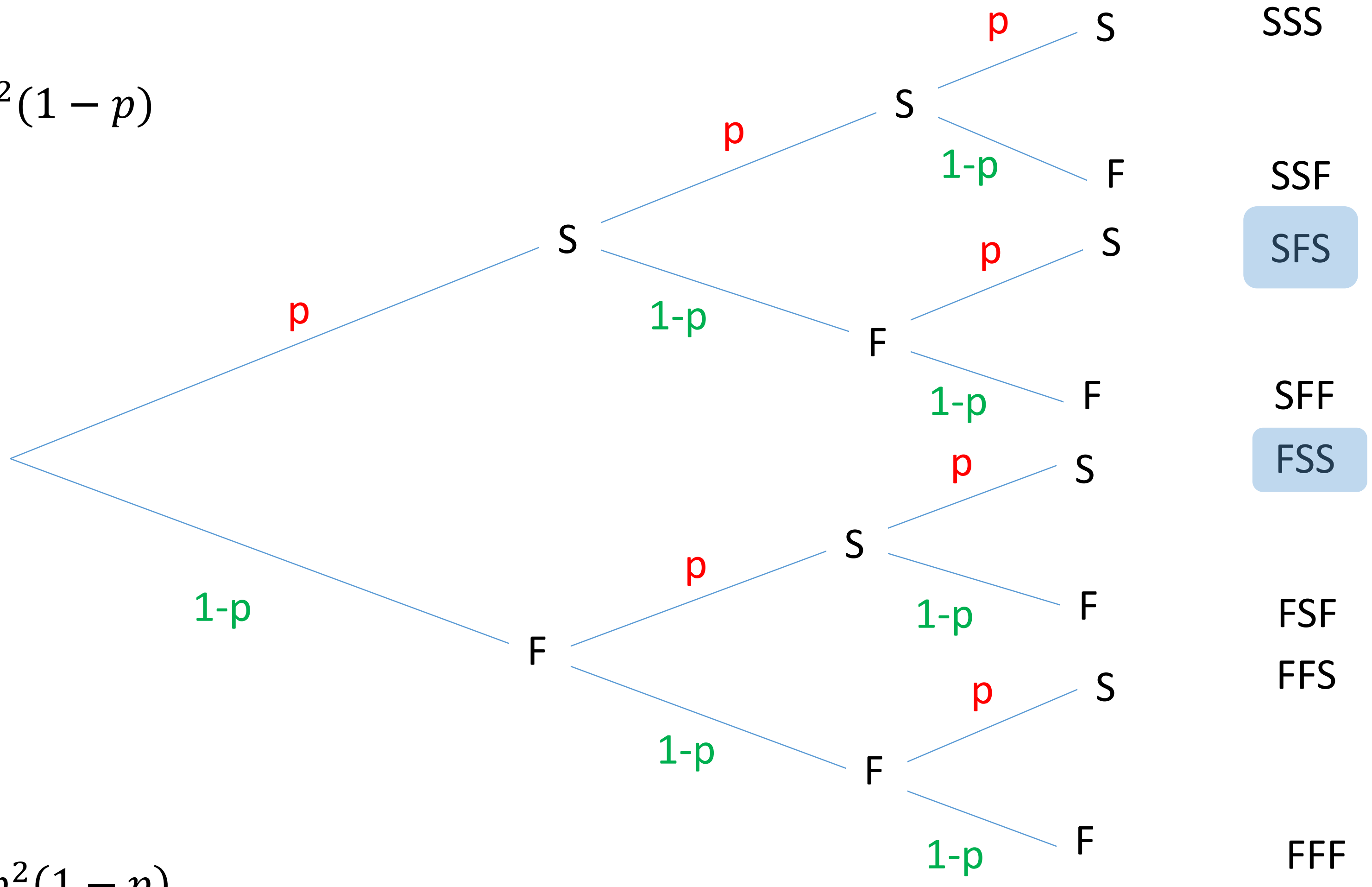
By definition, we know that the last trial is a success, thus we only have to see in how many ways can we have remaining $r - 1$ successes in the remaining $x - 1$ trials, knowing that order does not matter in these $x - 1$ trials.

$X = \boldsymbol{number\ of\ trials}\ till\ \boldsymbol{r}\ successes$

What is the probability that it takes three trials (not more or less) to get first two successes?

$$P(S \cap F \cap S) + P(F \cap S \cap S)$$

$$= p(1-p)p + (1-p)pp \;\; = 2p^2(1-p)$$



Using the Negative Binomial Distribution formula with $x = 3, r = 2$ we get the same result

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r$$

$$f(3) = \binom{3-1}{2-1}p^2(1-p) = \frac{2!}{(2-1)!\,1!}p^2(1-p)$$
$$= 2p^2(1-p)$$

Summary of Properties (Range, Mean, Variance)

Binomial Distribution

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$x = 0, 1, \ldots, n$$

$$\mu_X = E[X] = np$$

$$\sigma_X^2 = V[X] = np(1-p)$$

Geometric Distribution

$$f(x) = (1-p)^{x-1} p$$

$$x = 1, 2, \ldots$$

$$\mu_X = E[X] = \frac{1}{p}$$

$$\sigma_X^2 = V[X] = \frac{1-p}{p^2}$$

Negative Binomial Distribution

$$f(x) = \binom{x-1}{r-1} (1-p)^{x-r} p^r$$

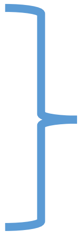
$$x = r, r+1, r+2, \ldots$$

$$\mu_X = E[X] = \frac{r(1-p)}{p}$$

$$\sigma_X^2 = V[X] = \frac{r(1-p)}{p^2}$$

What are some of the common distribution functions for discrete random variables?

- Uniform Distribution
- Binomial Distribution
- Geometric Distribution
- Negative Binomial Distribution
- Poisson Distribution



Based on **Bernoulli Trials**

What are some of the common distribution functions for discrete random variables?

- Uniform Distribution
- Binomial Distribution
- Geometric Distribution
- Negative Binomial Distribution
- Poisson Distribution

Based on **Bernoulli Trials**

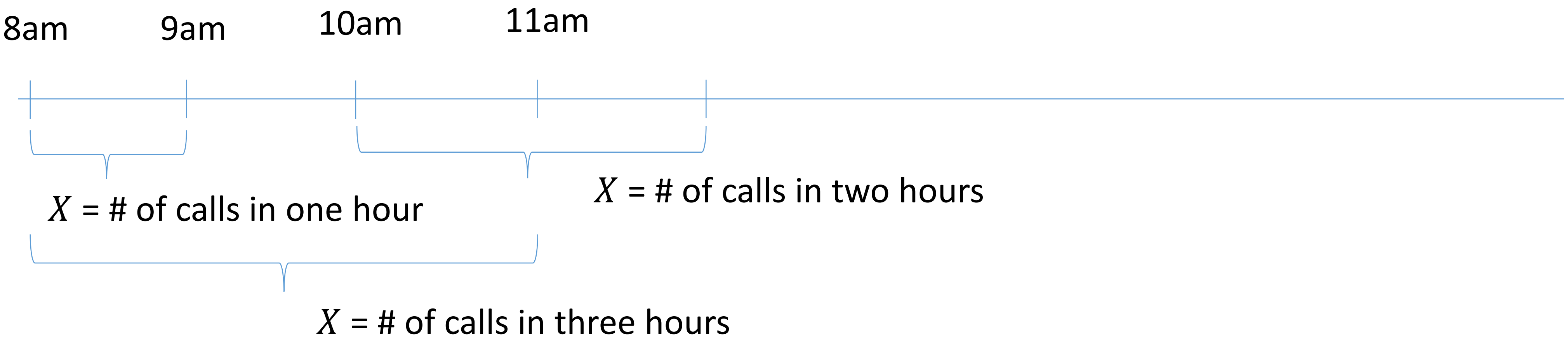
Based on a special "**counting**" process called **Poisson Process**

A Counting Process

- We often need to count random events, e.g.,
  - How many customers visit a shop in a given time duration
  - How many of my students fall asleep during my lecture
  - How many people visit a given website in a given time duration
  - How many calls you get during the lecture
  - How many drops of rain fall in a given area

A Counting Process

- Let X be the number of some specific random events that occur in duration, say t (t can be any unit of time, length, area, volume etc.)
  - E.g., if $t = 2$ hours, then X = {number of calls you get in two hours}
- Assume that the counting begins with zero, i.e., $X = 0$ for $t = 0$



8am 9am 10am 11am

X = # of calls in one hour

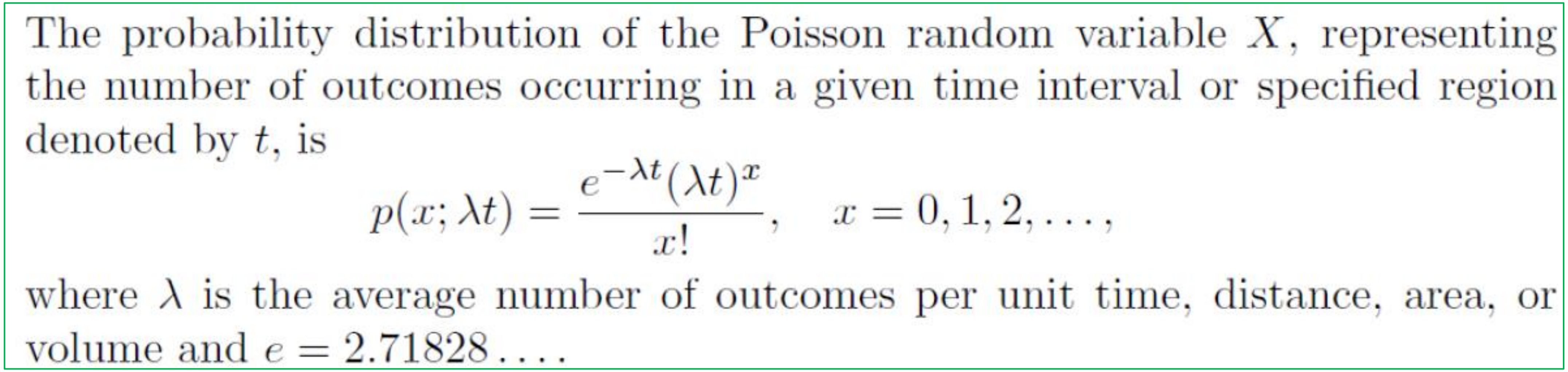
X = # of calls in two hours

X = # of calls in three hours

Poisson Process

- Let X be the number of some specific random events that occur in a duration, say t
- We say that X is a Poisson Random Variable if it satisfies the conditions of a Poisson Process, which are as follows
  - The number of events in two non-overlapping intervals are independent
  - The number of events during an interval depend only on the length of the interval
  - The probability of two events occurring at exactly the same time is zero
  - The average number of events per unit interval (time, distance etc.) is constant (usually denoted by λ)
    - Thus average number of events in interval t is λt

Poisson Distribution



The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region denoted by t, is

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots,$$

where λ is the average number of outcomes per unit time, distance, area, or volume and $e = 2.71828\ldots$.

Examples

Questions?? Thoughts??

Continuous Random Variables and Probability Distributions

3-1 CONTINUOUS RANDOM VARIABLES

A continuous random variable has an infinite number of possible values & the probability of any one particular value is zero. A *continuous random variable* is a random variable that can assume any value in some interval of numbers, and are thus NOT countable.

Examples:

- ✓ The time that a train arrives at a specified stop
- ✓ The lifetime of a transistor
- ✓ A randomly selected number between 0 and 1
- ✓ Let R be a future value of a weekly ratio of closing prices for IBM stock
- ✓ Let W be the exact weight of a randomly selected student

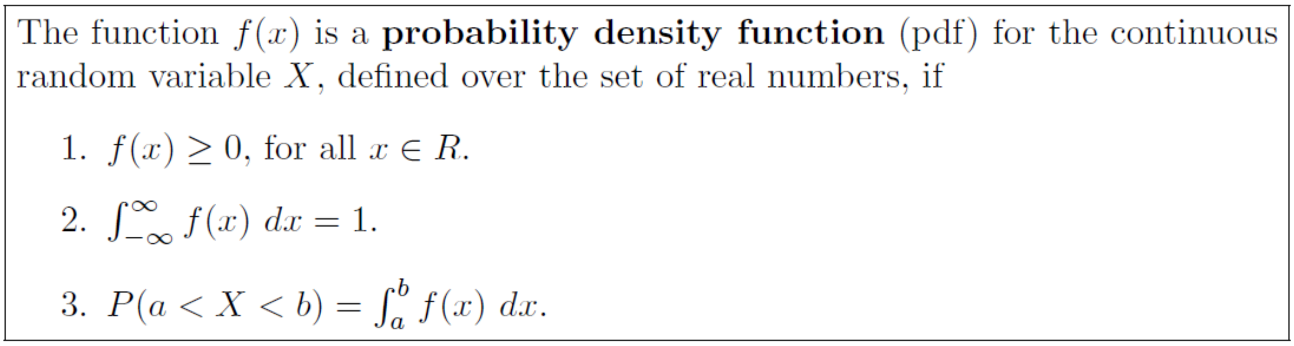
3-2 PROBABILITY DISTRIBUTIONS AND PROBABILITY DENSITY FUNCTIONS

A random variable is said to be continuous if there is a function $f(x)$ with the following properties:

a) Domain: all real numbers
b) Range: $f(x) \geq 0$
c) The area under the entire curve is 1

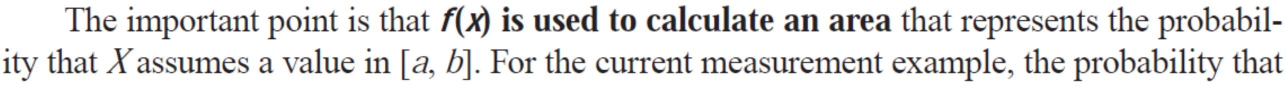
Such a function $f(x)$ is called the *probability density function* (abbreviated p.d.f.)

The fact that the total area under the curve $f(x)$ is 1 for all X values of the random variable tells us that all probabilities are expressed in terms of the area under the curve of this function.

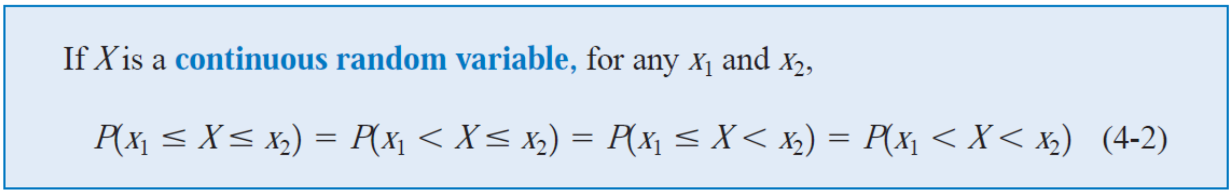


The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.



The important point is that **$f(x)$ is used to calculate an area** that represents the probability that X assumes a value in $[a, b]$. For the current measurement example, the probability that



If X is a **continuous random variable,** for any x_1 and x_2,

$$P(x_1 \leq X \leq x_2) = P(x_1 < X \leq x_2) = P(x_1 \leq X < x_2) = P(x_1 < X < x_2) \quad (4\text{-}2)$$

3-3 CUMULATIVE DISTRIBUTION FUNCTIONS

The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt, \quad \text{for } -\infty < x < \infty.$$

- **$0 \leq F(x) \leq 1$, for all x**

Notice that in the definition of $F(x)$, any $<$ can be changed to $\leq$ and vice versa. That is, $F(x)$ can be defined as either $0.05x$ or 0 at the end-point $x = 0$, and $F(x)$ can be defined as either $0.05x$ or 1 at the end-point $x = 20$. In other words, $F(x)$ is a continuous function. For a discrete random variable, $F(x)$ is not a continuous function. Sometimes, a continuous random variable is defined as one that has a continuous cumulative distribution function.

The probability density function of a continuous random variable can be determined from the cumulative distribution function by differentiating. Recall that the fundamental theorem of calculus states that

$$\frac{d}{dx}\int_{-\infty}^{x} f(u)\, du = f(x)$$

As an immediate consequence of Definition 3.7, one can write the two results

$$P(a < X < b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx},$$

if the derivative exists.

3-4 MEAN AND VARIANCE OF A CONTINUOUS RANDOM VARIABLE

Let X be a random variable with probability distribution $f(x)$. The **mean**, or **expected value**, of X is

$$\mu = E(X) = \sum_x x f(x)$$

if X is discrete, and

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$

if X is continuous.

Let X be a random variable with probability distribution $f(x)$ and mean μ. The variance of X is

$$\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 f(x), \qquad \text{if } X \text{ is discrete, and}$$

$$\sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx, \qquad \text{if } X \text{ is continuous.}$$

The positive square root of the variance, σ, is called the **standard deviation** of X.

The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

3.5 TWO OR MORE RANDOM VARIABLES

Joint Probability Distributions

The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y)\, dx\, dy$, for any region A in the xy plane.

Given the joint probability distribution $f(x, y)$ of the discrete random variables X and Y, the probability distribution $g(x)$ of X alone is obtained by summing $f(x, y)$ over the values of Y. Similarly, the probability distribution $h(y)$ of Y alone is obtained by summing $f(x, y)$ over the values of X. We define $g(x)$ and $h(y)$ to be the **marginal distributions** of X and Y, respectively. When X and Y are continuous random variables, summations are replaced by integrals. We can now make the following general definition.

The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_y f(x, y) \quad \text{and} \quad h(y) = \sum_x f(x, y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

for the continuous case.

If we write the marginal $f_X(x)$ in terms of the joint density, then this becomes

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x\, f_{X,Y}(x, y)\, dxdy$$

Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \text{ provided } h(y) > 0.$$

$$P(a < X < b \mid Y = y) = \int_a^b f(x|y)\,dx.$$

Definition 4. *Let X, Y be random variables (discrete or continuous). joint (cumulative) distribution function is*

$$F_{X,Y}(x,y) = \mathbf{P}(X \leq x, Y \leq y)$$

If X and Y are jointly continuous then we can compute the joint cdf from their joint pdf:

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \left[\int_{-\infty}^{y} f(u,v)\,dv \right] du$$

If we know the joint cdf, then we can compute the joint pdf by taking partial derivatives of the above :

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = f(x,y)$$

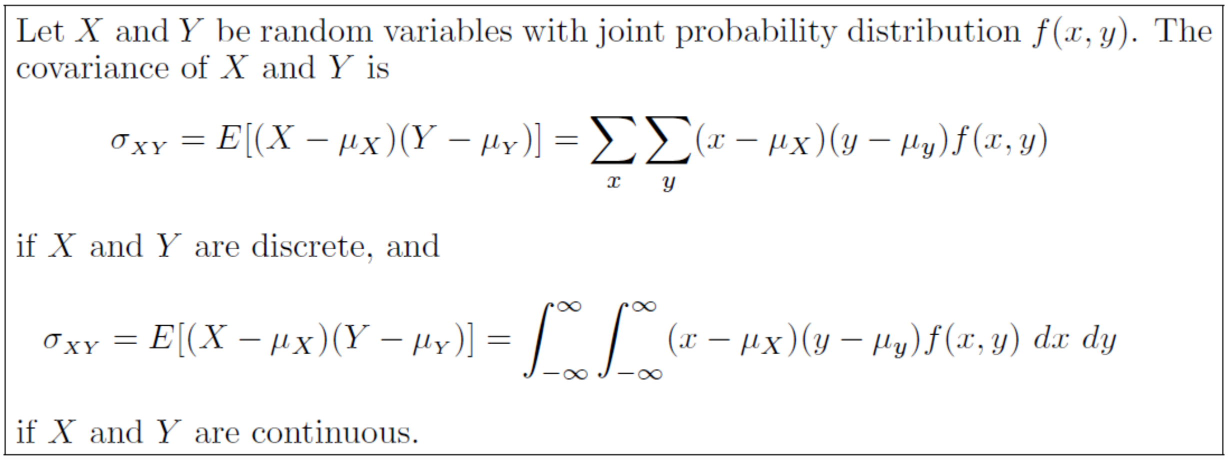
Statistical Independence

It should make sense to the reader that if $f(x|y)$ does not depend on y, then of course the outcome of the random variable Y has no impact on the outcome of the random variable X. In other words, we say that X and Y are independent random variables. We now offer the following formal definition of statistical independence.

Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x,y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x,y) = g(x)h(y)$$

for all (x,y) within their range.



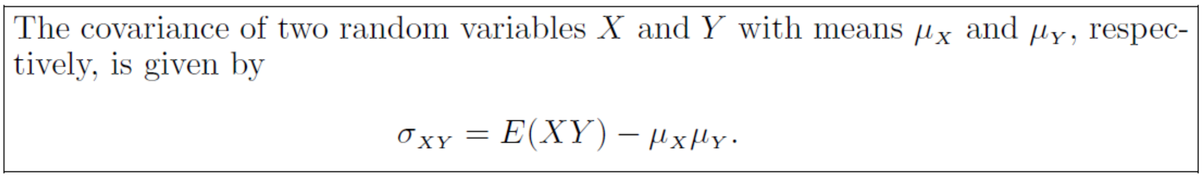
Let X and Y be random variables with joint probability distribution $f(x, y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x \sum_y (x - \mu_X)(y - \mu_y) f(x, y)$$

if X and Y are discrete, and

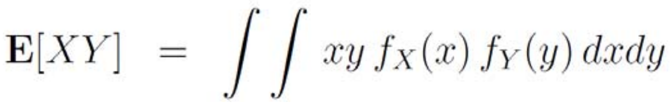
$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_y) f(x, y)\, dx\, dy$$

if X and Y are continuous.

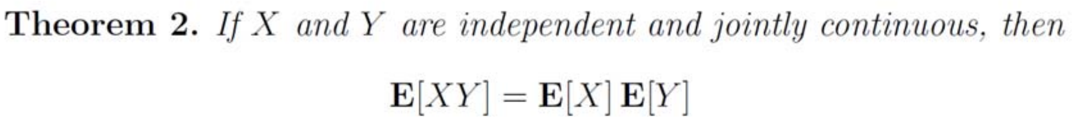


The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y.$$



$$\mathbf{E}[XY] \;=\; \int\int xy\, f_X(x)\, f_Y(y)\, dxdy$$



Theorem 2. *If X and Y are independent and jointly continuous, then*

$$\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$$

Continuous Random Variables and Probability Distributions

3-6 CONTINUOUS UNIFORM DISTRIBUTION

One of the simplest continuous distributions in all of statistics is the **continuous uniform distribution**. This distribution is characterized by a density function that is "flat," and thus the probability is uniform in a closed interval, say $[A, B]$. Although applications of the continuous uniform distribution are not as abundant as those for other distributions discussed in this chapter, it is appropriate for the novice to begin this introduction to continuous distributions with the uniform distribution.

Uniform Distribution

The density function of the continuous uniform random variable X on the interval $[A, B]$ is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A}, & A \le x \le B, \\ 0, & \text{elsewhere.} \end{cases}$$

The density function forms a rectangle with base $B-A$ and **constant height** $\frac{1}{B-A}$. As a result, the uniform distribution is often called the **rectangular distribution**. Note, however, that the interval may not always be closed: $[A, B]$. It can be (A, B) as well. The density function for a uniform random variable on the interval $[1, 3]$ is shown in Figure 6.1.

$$f(x) = \frac{1}{b-a} \qquad \text{if } a \le x \le b$$

$$f(x) = 0 \qquad \text{otherwise}$$

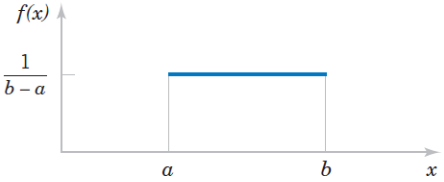


Figure 4-8 Continuous uniform probability density function.

The mean and variance of the uniform distribution are

$$\mu = \frac{A+B}{2} \text{ and } \sigma^2 = \frac{(B-A)^2}{12}.$$

Proof

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{0.5x^2}{b-a}\Big|_a^b = \frac{(a+b)}{2}$$

The variance of X is

$$V(X) = \int_a^b \frac{\left(x - \left(\frac{a+b}{2}\right)\right)^2}{b-a}\,dx = \frac{\left(x - \frac{a+b}{2}\right)^3}{3(b-a)}\Bigg|_a^b = \frac{(b-a)^2}{12}$$

The cumulative distribution function of a continuous uniform random variable is obtained by integration. If $a < x < b$,

$$F(x) = \int_{a}^{x} 1/(b - a)\,du = x/(b - a) - a/(b - a)$$

Therefore, the complete description of the cumulative distribution function of a continuous uniform random variable is

$$F(x) = \begin{cases} 0 & x < a \\ (x - a)/(b - a) & a \leq x < b \\ 1 & b \leq x \end{cases}$$

3-7 NORMAL DISTRIBUTION

Undoubtedly, the most widely used model for the distribution of a random variable is a **normal distribution.** Whenever a random experiment is replicated, the random variable that equals the average (or total) result over the replicates tends to have a normal distribution as the number of replicates becomes large. De Moivre presented this fundamental result, known as the **central limit theorem,** in 1733. Unfortunately, his work was lost for some time, and Gauss independently developed a normal distribution nearly 100 years later. Although De Moivre was later credited with the derivation, a normal distribution is also referred to as a **Gaussian distribution.**

The theoretical basis of a normal distribution is mentioned to justify the somewhat complex form of the probability density function. Our objective now is to calculate probabilities for a normal random variable. The central limit theorem will be stated more carefully later.

Random variables with different means and variances can be modeled by normal probability density functions with appropriate choices of the center and width of the curve. The value of $E(X) = \mu$ determines the center of the probability density function and the value of $V(X) = \sigma^2$ determines the width. Figure 4-10 illustrates several normal probability density functions with selected values of μ and σ^2. Each has the characteristic symmetric bell-shaped curve, but the centers and dispersions differ. The following definition provides the formula for normal probability density functions.

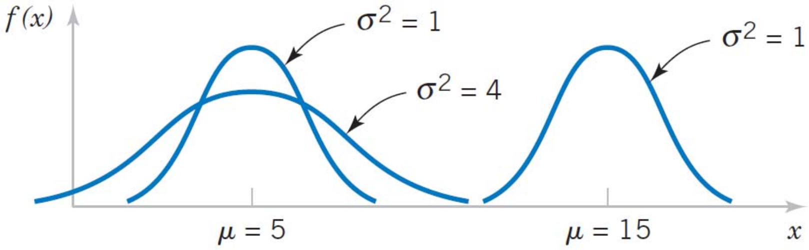


Figure 4-10 Normal probability density functions for selected values of the parameters μ and σ^2.

Normal Distribution

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{4-8}$$

is a **normal random variable** with parameters μ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \tag{4-9}$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution.

***Proof*:** To evaluate the mean, we first calculate

$$E(X-\mu) = \int_{-\infty}^{\infty} \frac{x-\mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx.$$

Setting $z = (x-\mu)/\sigma$ and $dx = \sigma\, dz$, we obtain

$$E(X-\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz = 0,$$

since the integrand above is an odd function of z. Using Theorem 4.5 on page 128, we conclude that

$$E(X) = \mu.$$

The variance of the normal distribution is given by

$$E[(X-\mu)^2] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}[(x-\mu)/\sigma]^2}\ dx.$$

Again setting $z = (x-\mu)/\sigma$ and $dx = \sigma\, dz$, we obtain

$$E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}}\ dz.$$

Integrating by parts with $u = z$ and $dv = ze^{-z^2/2}\ dz$ so that $du = dz$ and $v = -e^{-z^2/2}$, we find that

$$E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \left(-ze^{-z^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-z^2/2}\ dz \right) = \sigma^2(0+1) = \sigma^2.$$

Some useful results concerning a normal distribution are summarized below and in Fig. 4-12. For any normal random variable,

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$

Also, from the symmetry of $f(x)$, $P(X > \mu) = P(X < \mu) = 0.5$. Because $f(x)$ is positive for all x, this model assigns some probability to each interval of the real line. However, the

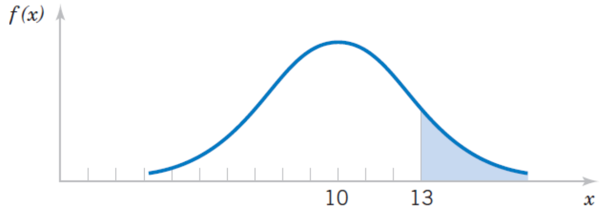


Figure 4-11 Probability that $X > 13$ for a normal random variable with $\mu = 10$ and $\sigma^2 = 4$.

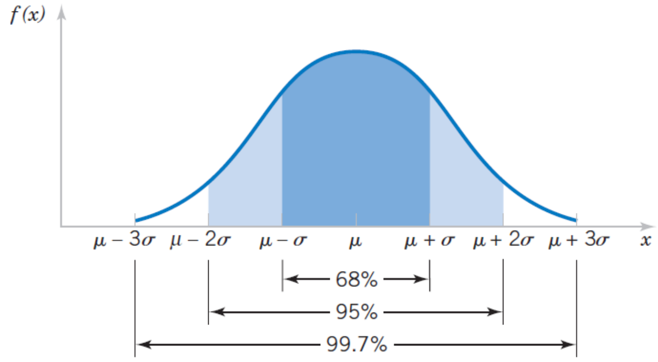


Figure 4-12 Probabilities associated with a normal distribution.

The Gaussian random variable is described in terms of two parameters $m \in \mathbb{R}$ and $\sigma > 0$ by the PDF

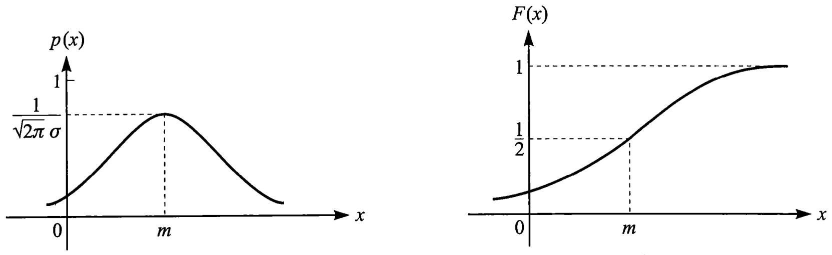
$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{2.3–8}$$

We usually use the shorthand form $\mathcal{N}(m, \sigma^2)$ to denote the PDF of Gaussian random variables and write $X \sim \mathcal{N}(m, \sigma^2)$. For this random variable

$$\begin{aligned} \mathrm{E}[X] &= m \\ \mathrm{VAR}[X] &= \sigma^2 \end{aligned} \tag{2.3–9}$$

A Gaussian random variable with $m = 0$ and $\sigma = 1$ is called a *standard normal.* A function closely related to the Gaussian random variable is the Q function defined as

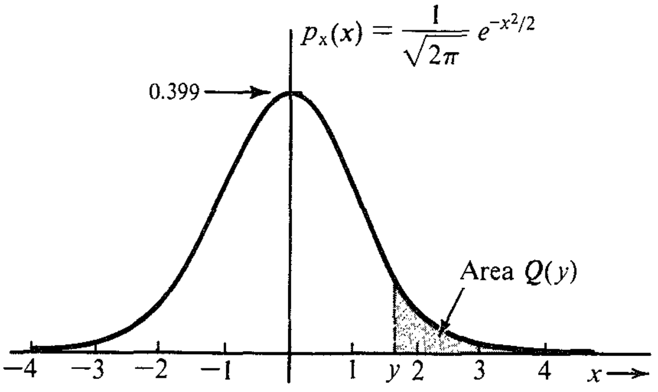
$$Q(x) = \mathrm{P}[\mathcal{N}(0, 1) > x] = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}}\, dt \tag{2.3–10}$$



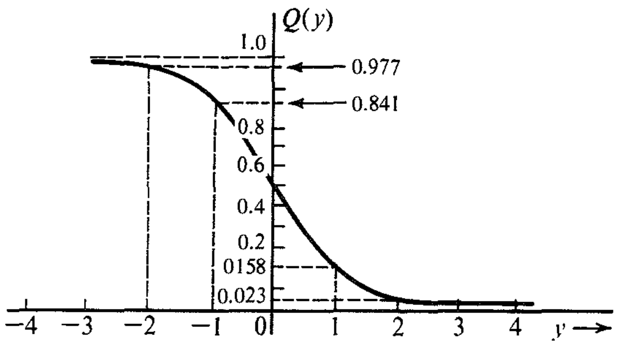
(a)

(b)

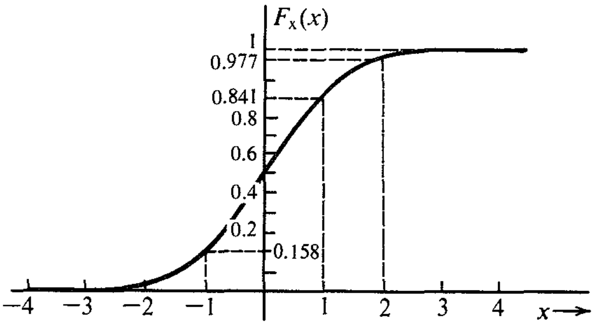
FIGURE 2.3–1
PDF and CDF of a Gaussian random variable.



(a)



(b)



(c)

The CDF of a Gaussian random variable is given by

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt \\ &= 1 - \int_{x}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt \\ &= 1 - \int_{\frac{x-m}{\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du \\ &= 1 - Q\left(\frac{x-m}{\sigma}\right) \end{aligned} \tag{2.3–11}$$

where we have introduced the change of variable $u = (t - m)/\sigma$. The PDF and the CDF of a Gaussian random variable are shown in Figure 2.3–1.

In general if $X \sim \mathcal{N}(m, \sigma^2)$, then

$$\begin{aligned} \mathrm{P}[X > \alpha] &= Q\left(\frac{\alpha - m}{\sigma}\right) \\ \mathrm{P}[X < \alpha] &= Q\left(\frac{m - \alpha}{\sigma}\right) \end{aligned} \tag{2.3–12}$$

Following are some of the important properties of the Q function:

$$Q(0) = \frac{1}{2} \qquad Q(\infty) = 0 \tag{2.3–13}$$

$$Q(-\infty) = 1 \qquad Q(-x) = 1 - Q(x) \tag{2.3–14}$$

$$Q(x) \approx \frac{1}{x\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{2.3–16}$$

The distribution of a normal random variable with mean 0 and variance 1 is called a **standard normal distribution**.

Creating a new random variable by this transformation is referred to as **standardizing.** The random variable Z represents the distance of X from its mean in terms of standard deviations. It is the key step to calculating a probability for an arbitrary normal random variable.

Standardizing to Calculate a Probability

Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P(Z \le z) \qquad (4\text{-}11)$$

where Z is a **standard normal random variable**, and $z = \frac{(x-\mu)}{\sigma}$ is the ***z*-value** obtained by **standardizing** X. The probability is obtained by using Appendix Table III with $z = (x-\mu)/\sigma$.

3-8 EXPONENTIAL DISTRIBUTION

Exponential Distribution

The random variable X that equals the distance between successive events of a Poisson process with mean number of events $\lambda > 0$ per unit interval is an **exponential random variable** with parameter λ. The probability density function of X is

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad 0 \le x < \infty \qquad (4\text{-}14)$$

Mean and Variance

If the random variable X has an exponential distribution with parameter λ,

$$\mu = E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = V(X) = \frac{1}{\lambda^2} \qquad (4\text{-}15)$$

3.9 The Central Limit Theorem

What it the central limit theorem? The theorem says that under rather general circumstances, if you sum *independent* random variables and normalize them accordingly, then at the limit (when you sum lots of them) you'll get a normal distribution.

For reference, here is the density of the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2:

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

We now state a very weak form of the central limit theorem. *Suppose that X_i are independent, identically distributed random variables with zero mean and variance σ^2. Then*

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \longrightarrow \mathcal{N}(0, \sigma^2).$$

As the sample size n increases, the distribution of the sample mean μ of a random sample from a population (not necessarily normal) with mean μ and variance σ^2 approaches normal with mean μ and variance σ^2/n.

Central Limit Theorem: Suppose $X_1, \ldots, X_n$ are i.i.d. with E$[X_i] = \mu$ and Var$(X_i) = \sigma^2$. Then as $n \to \infty$,

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} \text{Nor}(0,1),$$

where" $\xrightarrow{\mathcal{D}}$ " means that the c.d.f. → the Nor$(0,1)$ c.d.f.

Remarks: (1) So if n is large, then $\bar{X} \approx$ Nor$(\mu, \sigma^2/n)$.

(2) The X_i's *don't have to be normal* for the CLT to work!

(3) You usually need $n \geq 30$ observations for the approximation to work well. (Need fewer observations if the X_i's come from a symmetric distribution.)

(4) You can almost always use the CLT if the observations are i.i.d.

The normal approximation for $\bar{X}$ will generally be good if $n \geq 30$, provided the population distribution is not terribly skewed. If $n < 30$, the approximation is good only if the population is not too different from a normal distribution and, as stated above, if the population is known to be normal, the sampling distribution of $\bar{X}$ will follow a normal distribution exactly, no matter how small the size of the samples.

The sample size $n = 30$ is a guideline to use for the Central Limit Theorem. However, as the statement of the theorem implies, the presumption of normality on the distribution of $\bar{X}$ becomes more accurate as n grows larger. In fact, Figure 8.1 illustrates how the theorem works. It shows how the distribution of $\bar{X}$ becomes closer to normal as n grows larger, beginning with the clearly nonsymmetric distribution of an individual observation ($n = 1$). It also illustrates that the mean of $\bar{X}$ remains μ for any sample size and the variance of $\bar{X}$ gets smaller as n increases.

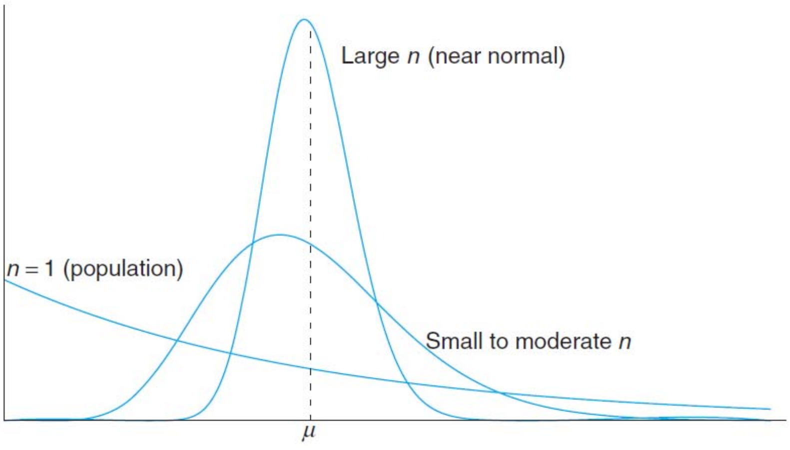


Figure 8.1: Illustration of the Central Limit Theorem (distribution of $\bar{X}$ for $n = 1$, moderate n, and large n).

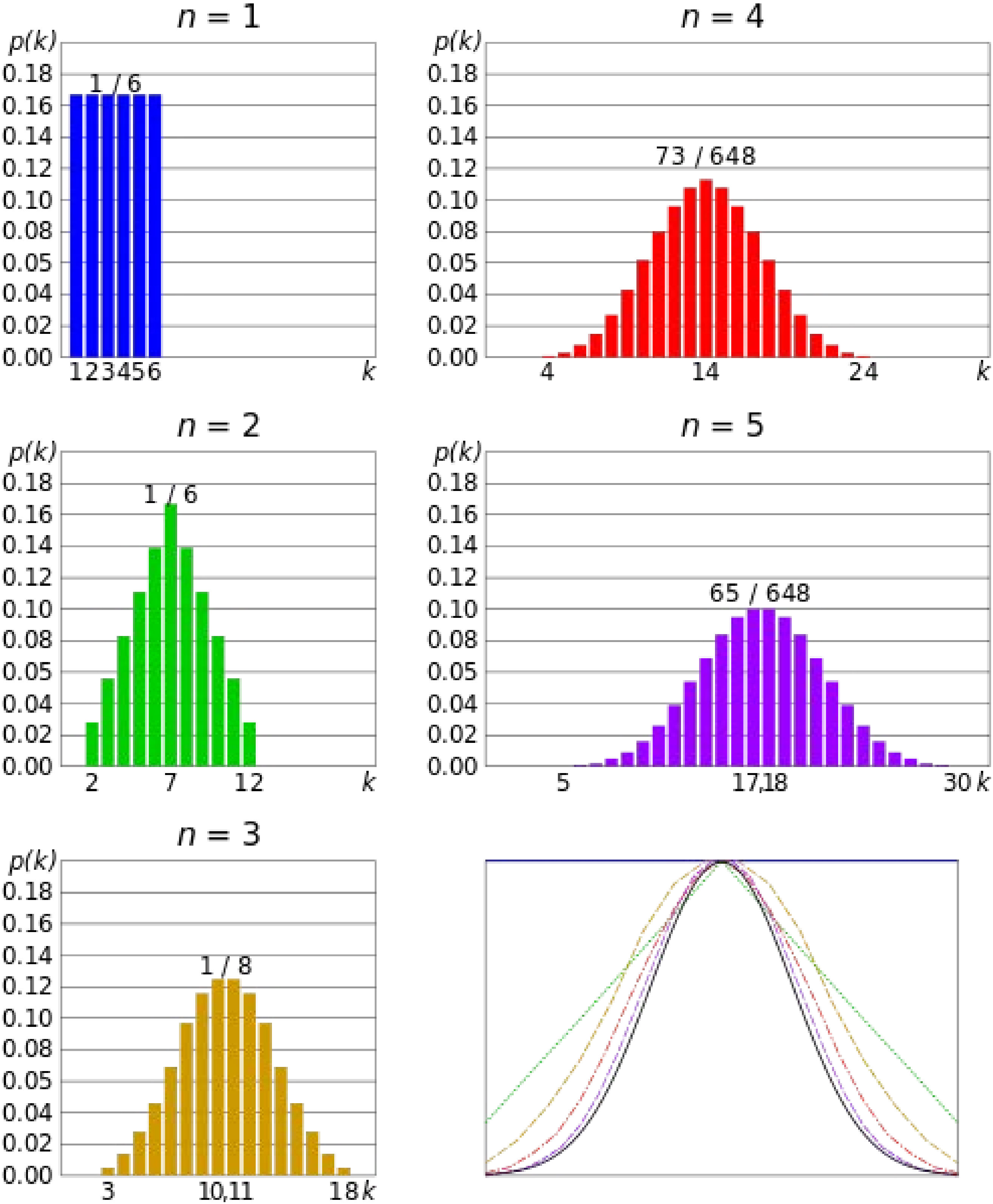

n = 1
p(k)
0.18
0.16
0.14
0.12
0.10
0.08
0.05
0.04
0.02
0.00
1 / 6
1 2 3 4 5 6
k
n = 4
p(k)
73 / 648
4
14
24
k
n = 2
p(k)
1 / 6
2
7
12
k
n = 5
p(k)
65 / 648
5
17,18
30
k
n = 3
p(k)
1 / 8
3
10,11
18
k

Table 1: Values of $Q(x)$ for $0 \leq x \leq 9$

x	$Q(x)$	x	$Q(x)$	x	$Q(x)$	x	$Q(x)$
0.00	0.5	2.30	0.010724	4.55	2.6823×10^{-6}	6.80	5.231×10^{-12}
0.05	0.48006	2.35	0.0093867	4.60	2.1125×10^{-6}	6.85	3.6925×10^{-12}
0.10	0.46017	2.40	0.0081975	4.65	1.6597×10^{-6}	6.90	2.6001×10^{-12}
0.15	0.44038	2.45	0.0071428	4.70	1.3008×10^{-6}	6.95	1.8264×10^{-12}
0.20	0.42074	2.50	0.0062097	4.75	1.0171×10^{-6}	7.00	1.2798×10^{-12}
0.25	0.40129	2.55	0.0053861	4.80	7.9333×10^{-7}	7.05	8.9459×10^{-13}
0.30	0.38209	2.60	0.0046612	4.85	6.1731×10^{-7}	7.10	6.2378×10^{-13}
0.35	0.36317	2.65	0.0040246	4.90	4.7918×10^{-7}	7.15	4.3389×10^{-13}
0.40	0.34458	2.70	0.003467	4.95	3.7107×10^{-7}	7.20	3.0106×10^{-13}
0.45	0.32636	2.75	0.0029798	5.00	2.8665×10^{-7}	7.25	2.0839×10^{-13}
0.50	0.30854	2.80	0.0025551	5.05	2.2091×10^{-7}	7.30	1.4388×10^{-13}
0.55	0.29116	2.85	0.002186	5.10	1.6983×10^{-7}	7.35	9.9103×10^{-14}
0.60	0.27425	2.90	0.0018658	5.15	1.3024×10^{-7}	7.40	6.8092×10^{-14}
0.65	0.25785	2.95	0.0015889	5.20	9.9644×10^{-8}	7.45	4.667×10^{-14}
0.70	0.24196	3.00	0.0013499	5.25	7.605×10^{-8}	7.50	3.1909×10^{-14}
0.75	0.22663	3.05	0.0011442	5.30	5.7901×10^{-8}	7.55	2.1763×10^{-14}
0.80	0.21186	3.10	0.0009676	5.35	4.3977×10^{-8}	7.60	1.4807×10^{-14}
0.85	0.19766	3.15	0.00081635	5.40	3.332×10^{-8}	7.65	1.0049×10^{-14}
0.90	0.18406	3.20	0.00068714	5.45	2.5185×10^{-8}	7.70	6.8033×10^{-15}
0.95	0.17106	3.25	0.00057703	5.50	1.899×10^{-8}	7.75	4.5946×10^{-15}
1.00	0.15866	3.30	0.00048342	5.55	1.4283×10^{-8}	7.80	3.0954×10^{-15}
1.05	0.14686	3.35	0.00040406	5.60	1.0718×10^{-8}	7.85	2.0802×10^{-15}
1.10	0.13567	3.40	0.00033693	5.65	8.0224×10^{-9}	7.90	1.3945×10^{-15}
1.15	0.12507	3.45	0.00028029	5.70	5.9904×10^{-9}	7.95	9.3256×10^{-16}
1.20	0.11507	3.50	0.00023263	5.75	4.4622×10^{-9}	8.00	6.221×10^{-16}
1.25	0.10565	3.55	0.00019262	5.80	3.3157×10^{-9}	8.05	4.1397×10^{-16}
1.30	0.0968	3.60	0.00015911	5.85	2.4579×10^{-9}	8.10	2.748×10^{-16}
1.35	0.088508	3.65	0.00013112	5.90	1.8175×10^{-9}	8.15	1.8196×10^{-16}
1.40	0.080757	3.70	0.0001078	5.95	1.3407×10^{-9}	8.20	1.2019×10^{-16}
1.45	0.073529	3.75	8.8417×10^{-5}	6.00	9.8659×10^{-10}	8.25	7.9197×10^{-17}
1.50	0.066807	3.80	7.2348×10^{-5}	6.05	7.2423×10^{-10}	8.30	5.2056×10^{-17}
1.55	0.060571	3.85	5.9059×10^{-5}	6.10	5.3034×10^{-10}	8.35	3.4131×10^{-17}
1.60	0.054799	3.90	4.8096×10^{-5}	6.15	3.8741×10^{-10}	8.40	2.2324×10^{-17}
1.65	0.049471	3.95	3.9076×10^{-5}	6.20	2.8232×10^{-10}	8.45	1.4565×10^{-17}
1.70	0.044565	4.00	3.1671×10^{-5}	6.25	2.0523×10^{-10}	8.50	9.4795×10^{-18}
1.75	0.040059	4.05	2.5609×10^{-5}	6.30	1.4882×10^{-10}	8.55	6.1544×10^{-18}
1.80	0.03593	4.10	2.0658×10^{-5}	6.35	1.0766×10^{-10}	8.60	3.9858×10^{-18}
1.85	0.032157	4.15	1.6624×10^{-5}	6.40	7.7688×10^{-11}	8.65	2.575×10^{-18}
1.90	0.028717	4.20	1.3346×10^{-5}	6.45	5.5925×10^{-11}	8.70	1.6594×10^{-18}
1.95	0.025588	4.25	1.0689×10^{-5}	6.50	4.016×10^{-11}	8.75	1.0668×10^{-18}
2.00	0.02275	4.30	8.5399×10^{-6}	6.55	2.8769×10^{-11}	8.80	6.8408×10^{-19}
2.05	0.020182	4.35	6.8069×10^{-6}	6.60	2.0558×10^{-11}	8.85	4.376×10^{-19}
2.10	0.017864	4.40	5.4125×10^{-6}	6.65	1.4655×10^{-11}	8.90	2.7923×10^{-19}
2.15	0.015778	4.45	4.2935×10^{-6}	6.70	1.0421×10^{-11}	8.95	1.7774×10^{-19}
2.20	0.013903	4.50	3.3977×10^{-6}	6.75	7.3923×10^{-12}	9.00	1.1286×10^{-19}
2.25	0.012224						

EE 302
Probabilistic Methods in Electrical Engineering

with

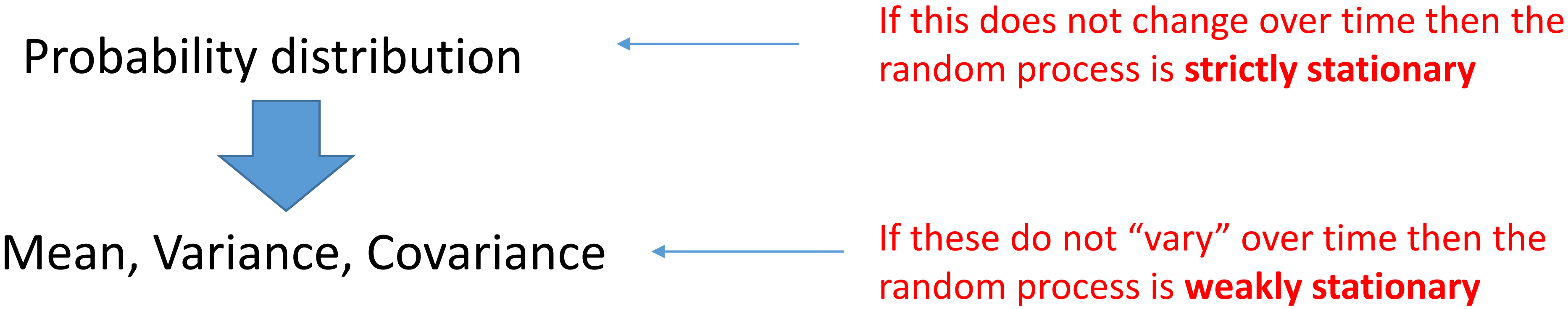
Dr. Naveed R. Butt

@

Jouf University

The Random and the Fixed

- Probability is a lack of knowledge.
- Randomly varying or uncertain events may still have some underlying characteristics that are "fixed"
- Some of these can be



Probability distribution

If this does not change over time then the random process is **strictly stationary**

Mean, Variance, Covariance

If these do not "vary" over time then the random process is **weakly stationary**

What is a random process?

- A random process is a **series of random events**
  - Can be discrete (student heights) or continuous (voice) or discretized (digitized voice)

Random Event/Variable

X

Height of next student who enters the room

Random Process (discrete)

$[X_1 \; X_2 \; X_3 \; X_4 \; X_5 \; X_6 \; X_7 \; X_8]$

Heights of next eight students who enter the room

But first …

- Let us be clear about what we mean by a random "process"
- A random process is a **series of random events**
  - Can be discrete (student heights) or continuous (voice) or discretized (discretized voice)

$X(t)$ *Human speech*

Random Process (continuous)

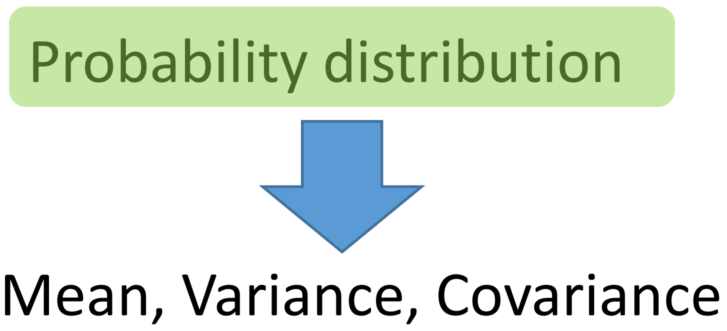
$[X(t_1)\ X(t_2)\ X(t_3)\ X(t_4)\ X(t_5)\ X(t_6)\ X(t_7)\ X(t_8)]$

Samples taken at times t1, t2, …, t8.

Random Process (discrete)

Our friends in an uncertain world!

- Random or uncertain events may still have some underlying characteristics that are “fixed”
- Some of these can be

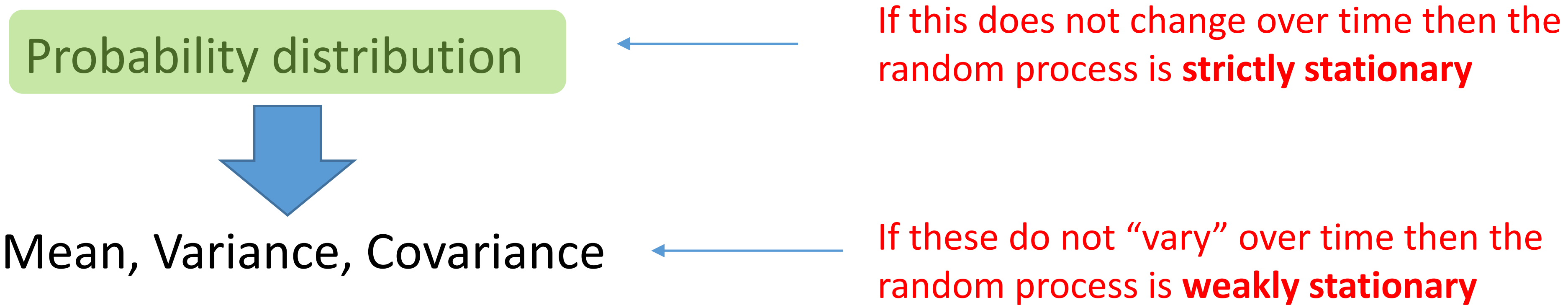


Probability distribution

Mean, Variance, Covariance

Our friends in an uncertain world!

- Random or uncertain events may still have some underlying characteristics that are “fixed”
- Some of these can be



Probability distribution

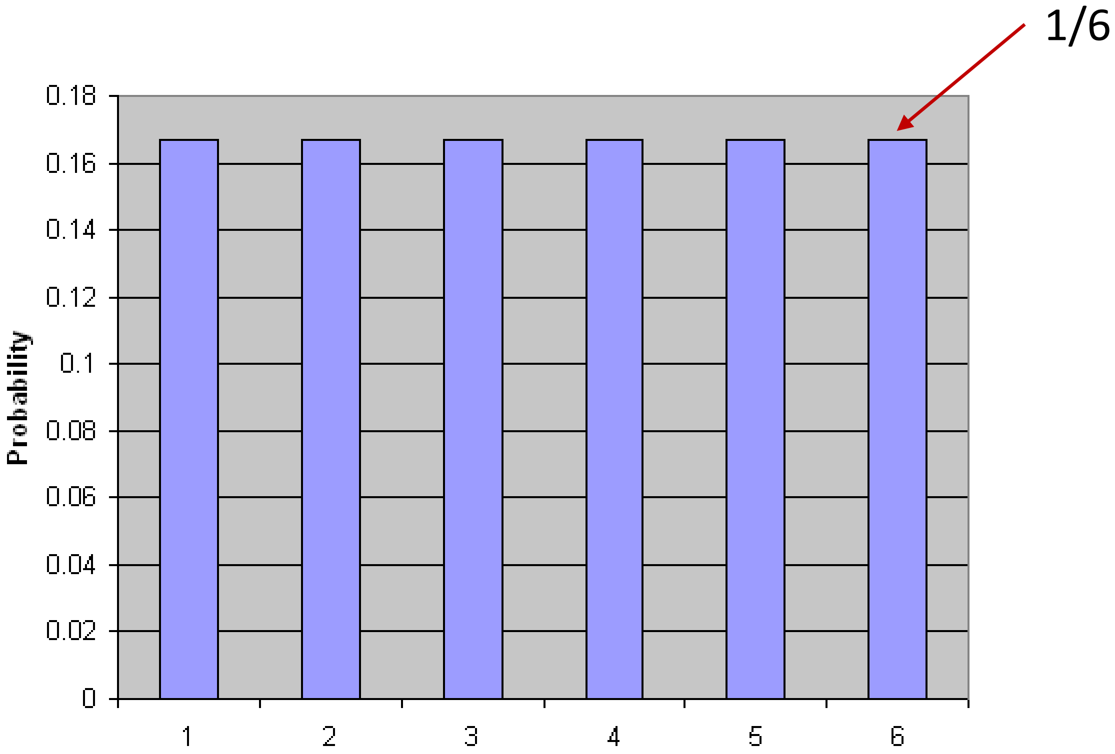
If this does not change over time then the random process is **strictly stationary**

Mean, Variance, Covariance

If these do not “vary” over time then the random process is **weakly stationary**

Revision - Distribution?

- A distribution is a collection of probabilities we assign to random events.
    - Collection? Graph, table, function



Our friends in an uncertain world!

- Random or uncertain events may still have some underlying characteristics that are "fixed"
- Some of these can be

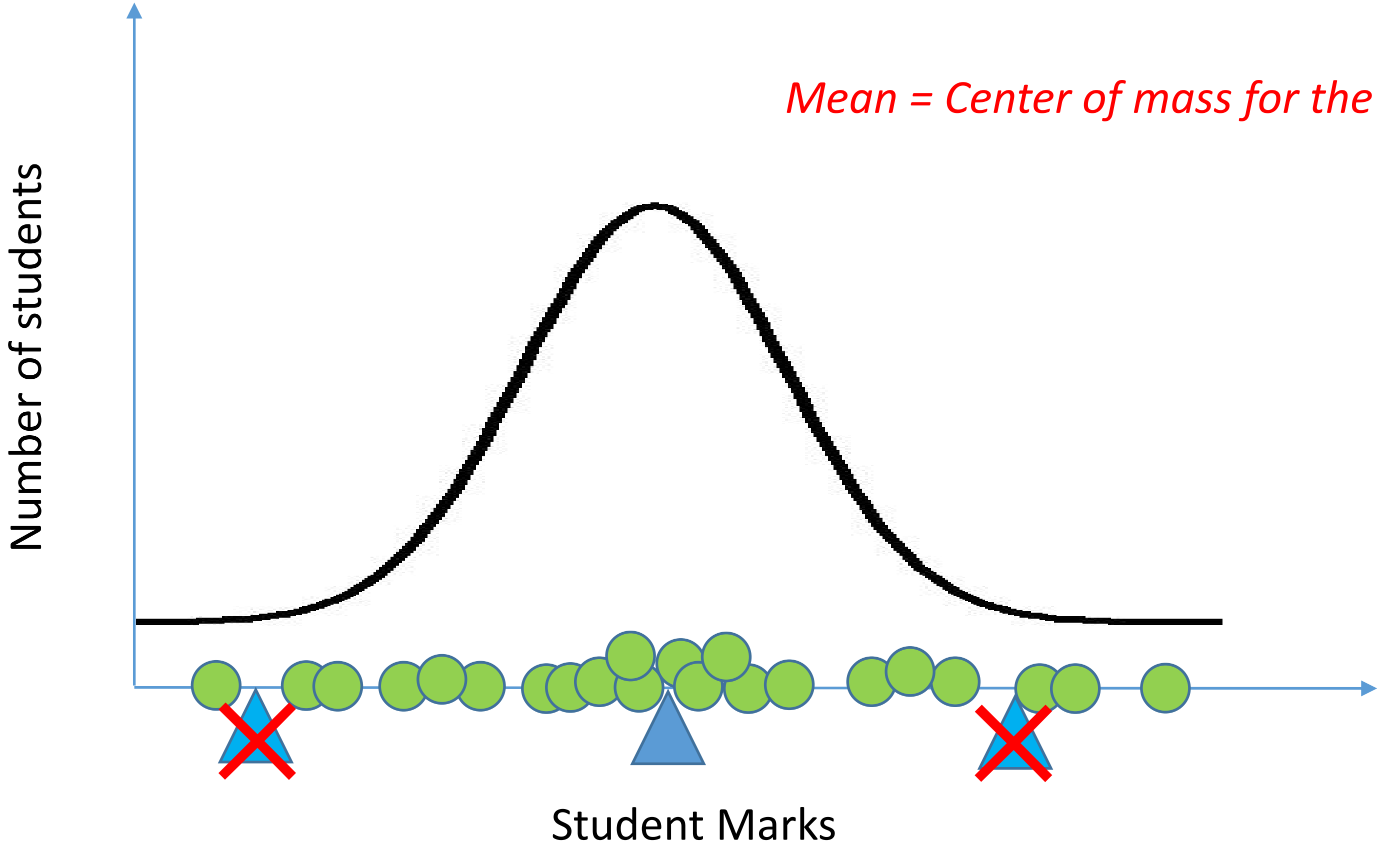
Probability distribution

If this does not change over time then the random process is **strictly stationary**

Mean, Variance, Covariance

If these do not "vary" over time then the random process is **weakly stationary**

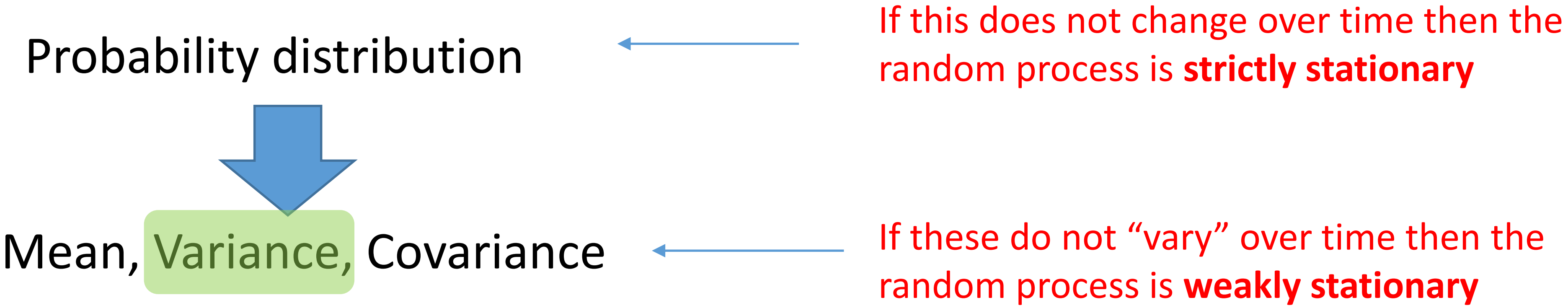
Revision: Mean/Average/Expectation



Mean = Center of mass for the distribution (balancing point)

Our friends in an uncertain world!

- Random or uncertain events may still have some underlying characteristics that are “fixed”
- Some of these can be



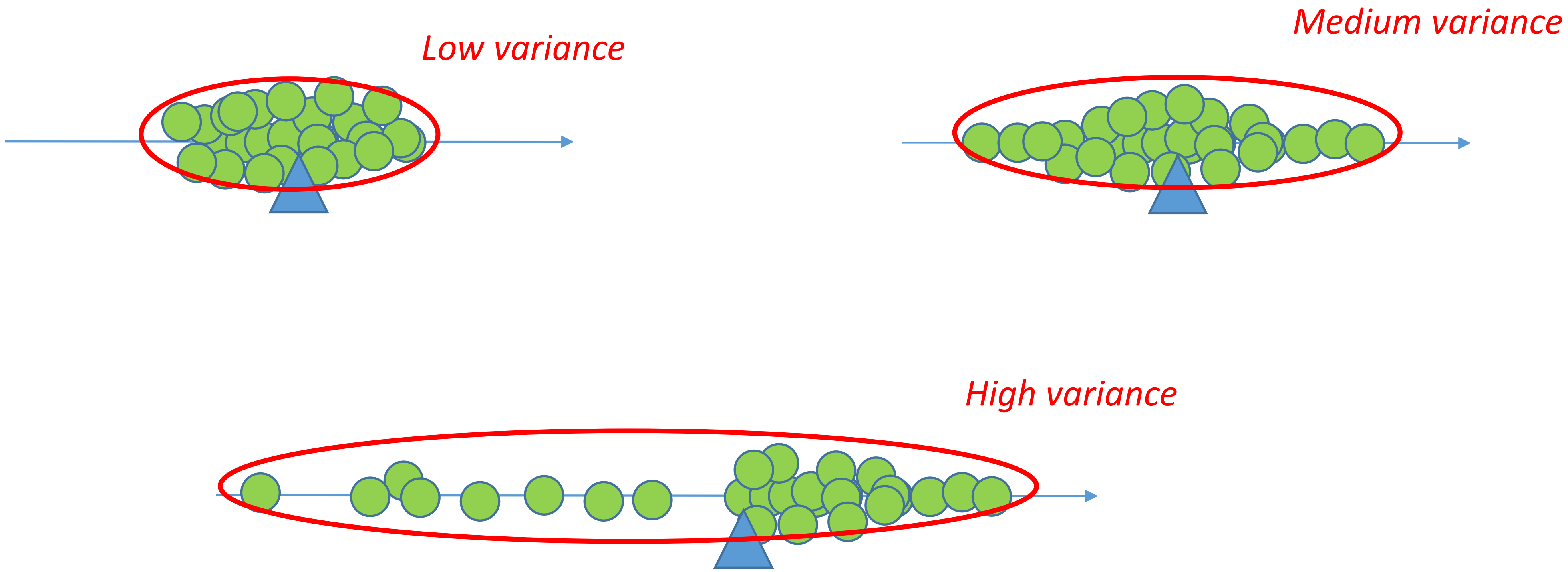
Probability distribution

Mean, Variance, Covariance

If this does not change over time then the random process is **strictly stationary**

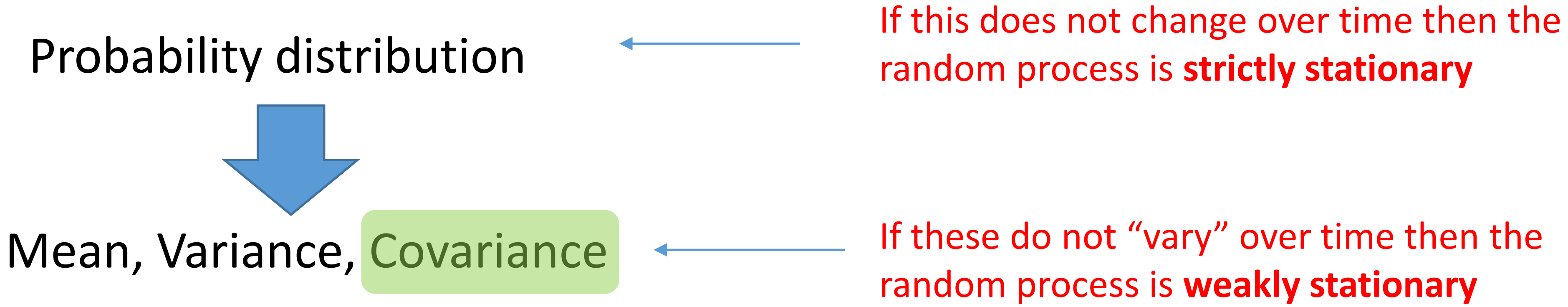
If these do not “vary” over time then the random process is **weakly stationary**

Revision: Variance = degree of spread (how much variation is there in the data/outcomes?)



Our friends in an uncertain world!

- Random or uncertain events may still have some underlying characteristics that are “fixed”
- Some of these can be



Probability distribution

Mean, Variance, Covariance

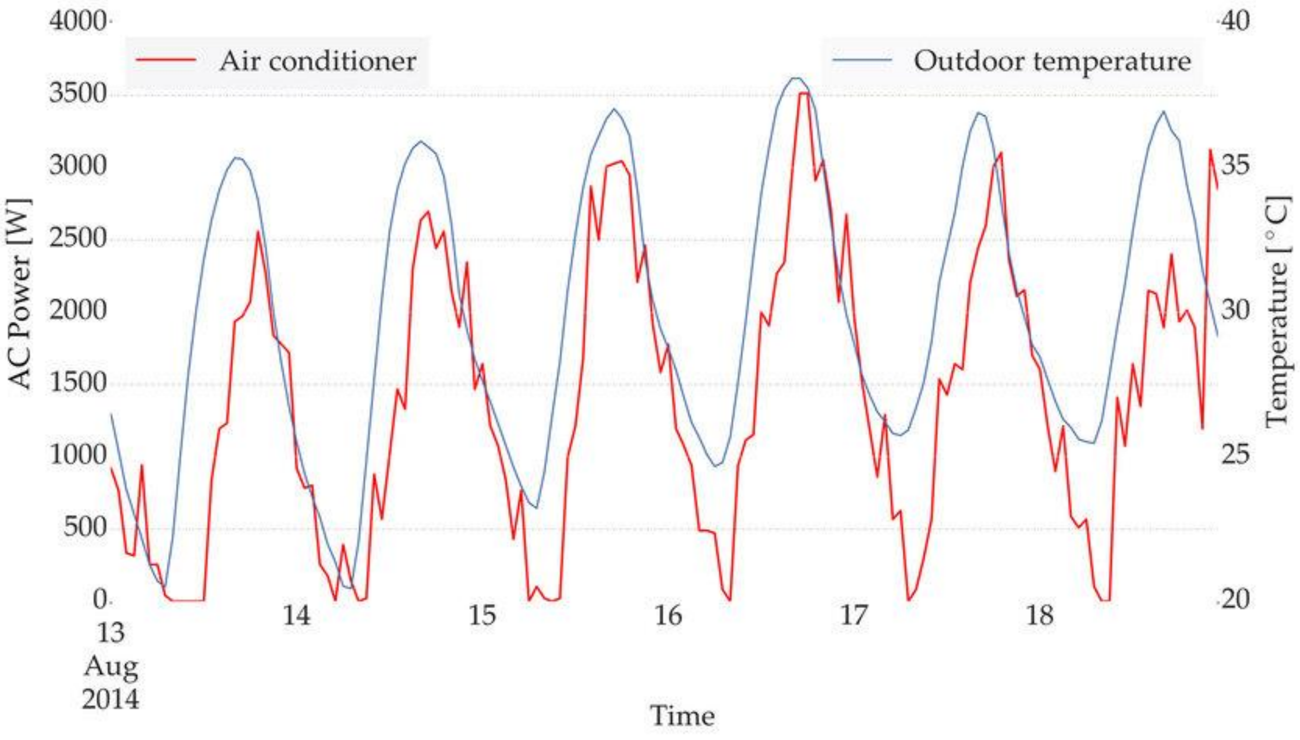
If this does not change over time then the random process is **strictly stationary**

If these do not “vary” over time then the random process is **weakly stationary**

Covariance = degree of linear relationship between data/outcomes

Can knowledge of one random process help us say something about the value of another?

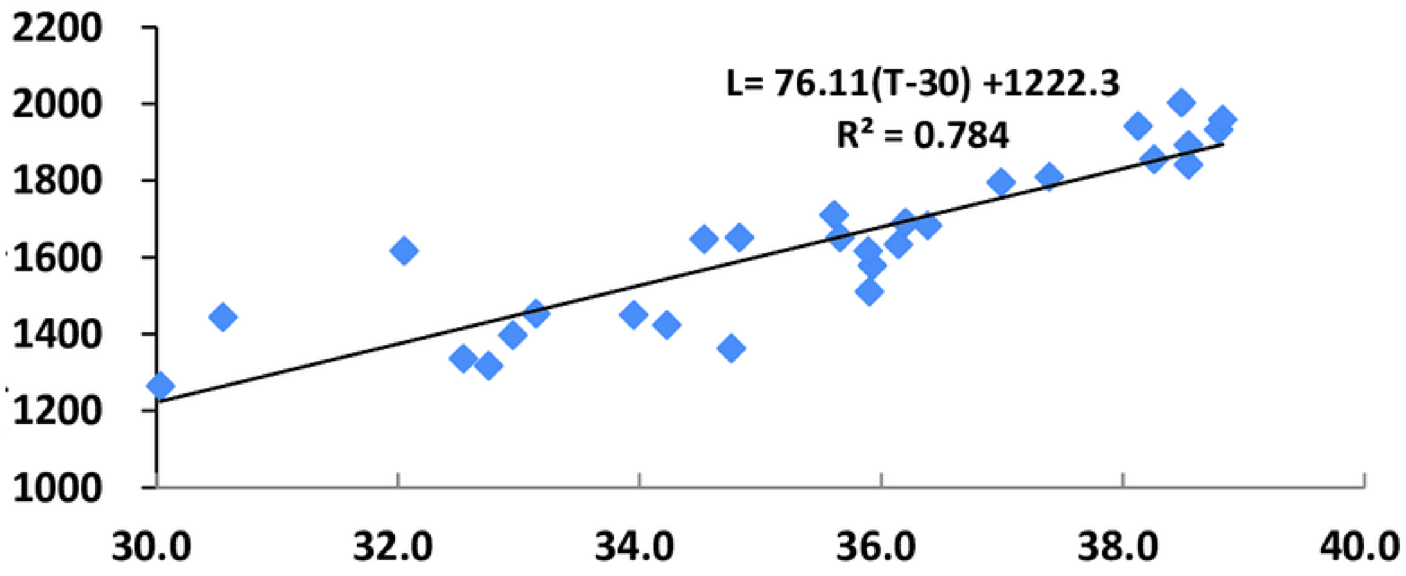
Yes – if their covariance is high!



Covariance = degree of linear relationship between data/outcomes

Can knowledge of one random process help us say something about the value of another?

Yes – if their covariance is high!

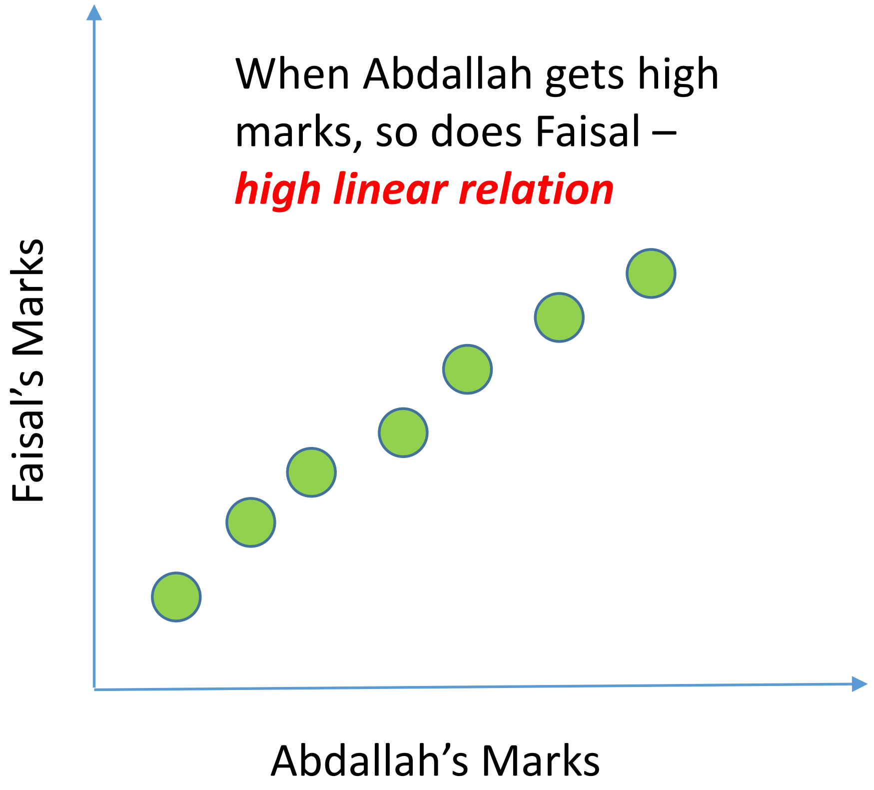


One way of checking linear relationship ("covariance") is to plot the two variables against each other.

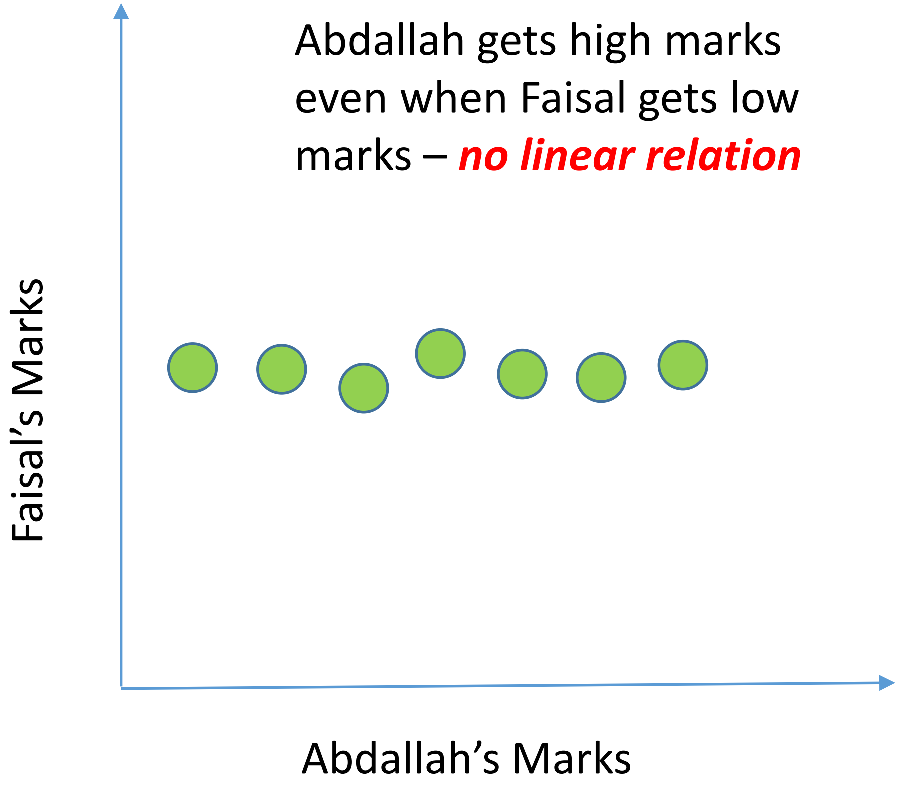
Covariance = degree of linear relationship between data/outcomes

	Abdallah’s Marks	**Faisal’s Marks**
Quiz 1	4	4
Quiz 2	8	7
Quiz 3	2	1
Quiz 4	4	5
Quiz 5	1	0
Quiz 6	9	10

Covariance = degree of linear relationship between data/outcomes

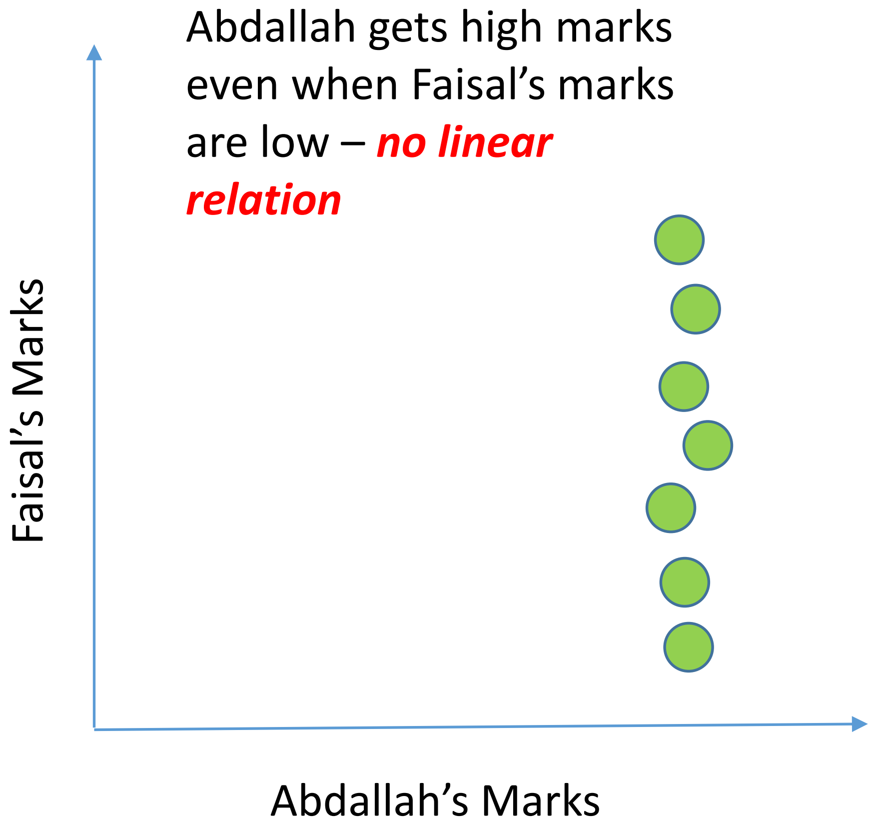


When Abdallah gets high marks, so does Faisal – ***high linear relation***

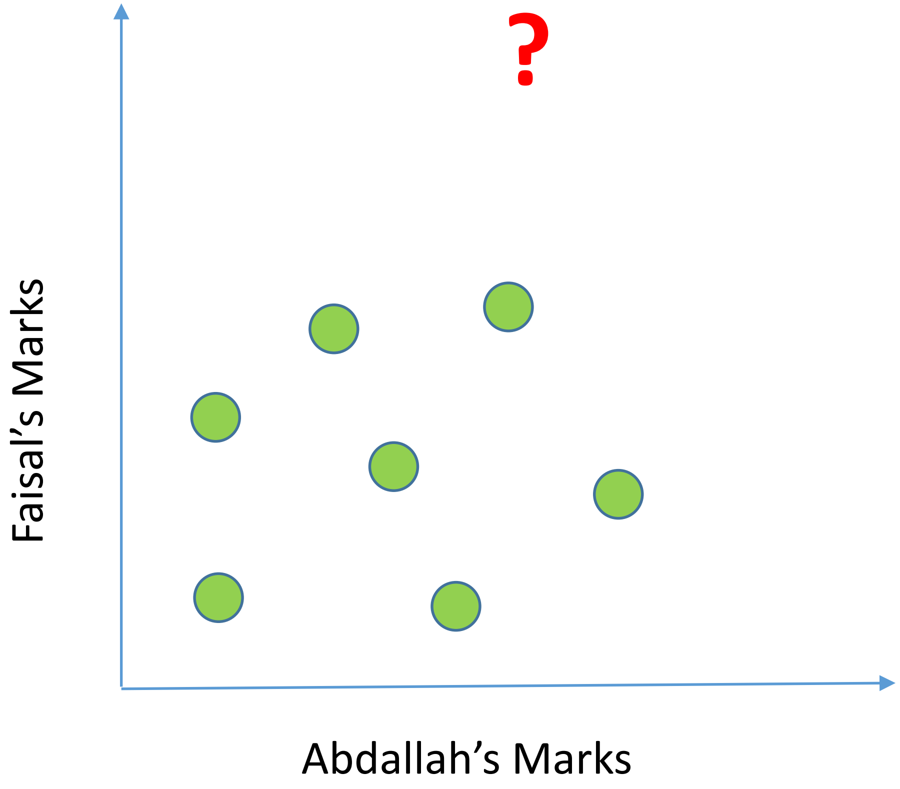


Abdallah gets high marks even when Faisal gets low marks – ***no linear relation***

Covariance = degree of linear relationship between data/outcomes



Abdallah gets high marks even when Faisal's marks are low – ***no linear relation***



Covariance = degree of linear relationship between data/outcomes

- Covariance helps us understand if two random processes are statistically related or not
  - **Statistically independent processes have zero covariance**
- We usually normalize covariance so that it lies between -1 and 1
  - Normalized covariance is called "**Correlation**"
- Noise in communications is mostly assumed to be independent of the message signal (i.e., we often assume zero correlation between signal and noise)

How do we write "covariance" mathematically?

Covariance of two random variables X and Y

$$C(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$
$$= \mathbb{E}[XY] - \mu_X \mu_Y$$

Independent random variables have zero covariance

$$C(X,Y) = 0$$

For a random process we use the ***autocovariance function***

$$C_{XX}(t_1, t_2) = \mathbb{E}[(X(t_1) - \mu_{X_{t_1}})(X(t_2) - \mu_{X_{t_2}})]$$

Covariance between two samples of random process X(t)

Mean and Autocovariance of a Weakly Stationary Process

The mean of a weakly stationary process does not change with time

$$\mu_X(t) = \mu_X \qquad \text{for all } t$$

The autocovariance function of a weakly stationary process depends only on the time difference between the samples and NOT on their actual values

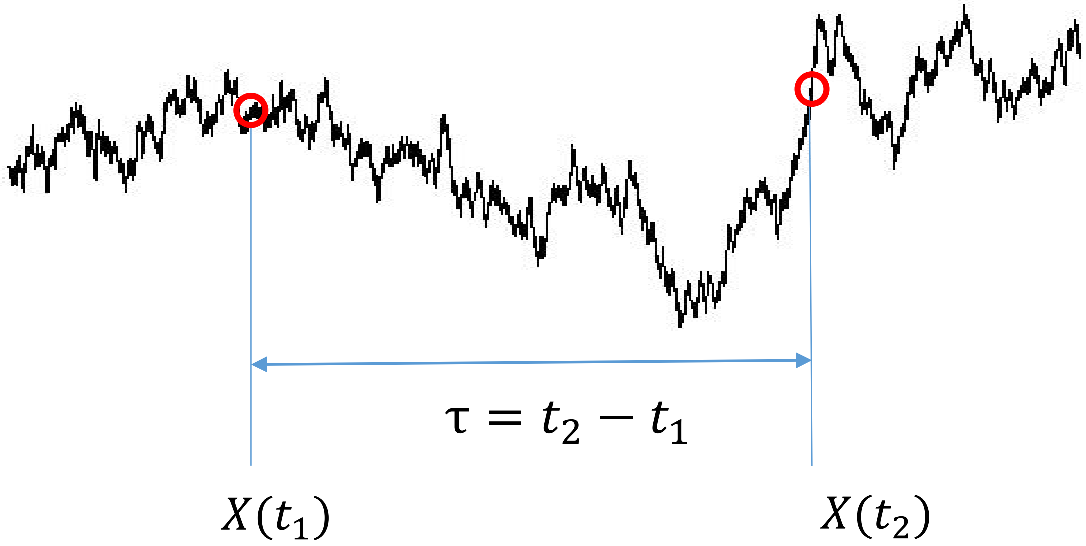
$$C_{XX}(t_2 - t_1) \qquad \text{for all } t_1 \text{ and } t_2$$

or

$$C_{XX}(\tau) \qquad \tau = t_2 - t_1$$

Calculating the autocorrelation function

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)]$$

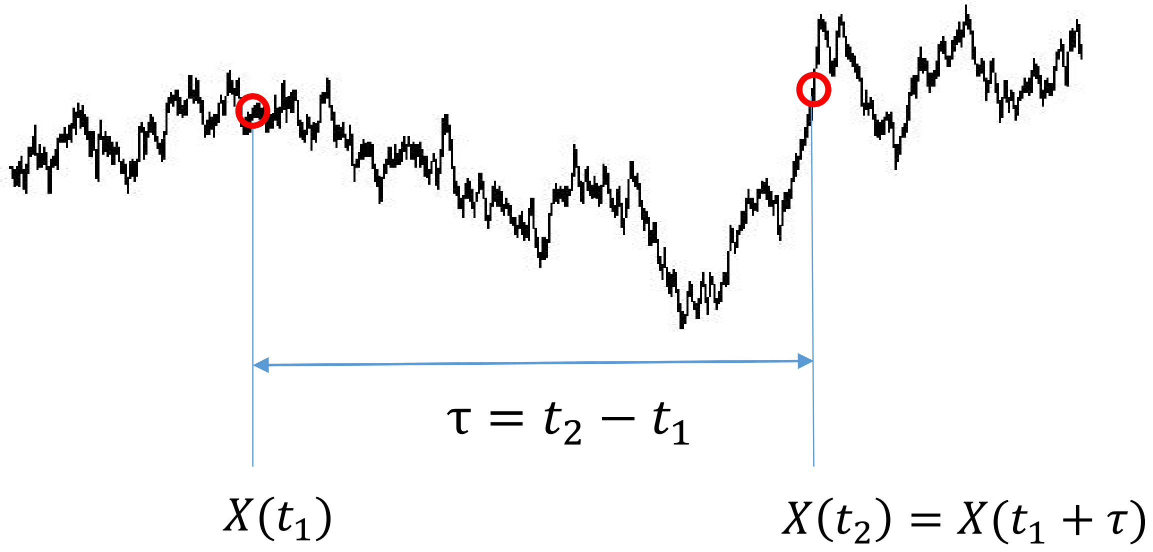


Calculating the autocorrelation function

- For a weakly stationary process X(t), we have

$$R_{XX}(t_1, t_2) = R_{XX}(t_2 - t_1) = R_{XX}(\tau)$$

$$R_{XX}(\tau) = \mathbb{E}[X(t+\tau)X(t)]$$



Mean and Autocorrelation of a Weakly Stationary Process

The mean of a weakly stationary process does not change with time

$$\mu_X(t) = \mu_X \qquad \text{for all } t$$

The autocorrelation function of a weakly stationary process depends only on the time difference between the samples and NOT on their actual values

if $t_2 - t_1 = t_4 - t_3$ then

$$R_{XX}(t_2 - t_1) = R_{XX}(t_4 - t_3)$$

Autocovariance vs. Autocorrelation of a weakly stationary process

Autocovariance function

$$C_{XX}(\tau) = \mathbb{E}[(X(t+\tau) - \mu_X)(X(t) - \mu_X)]$$

Covariance between two samples of random process X(t)

Only a minor difference between the two. Conceptually very similar.

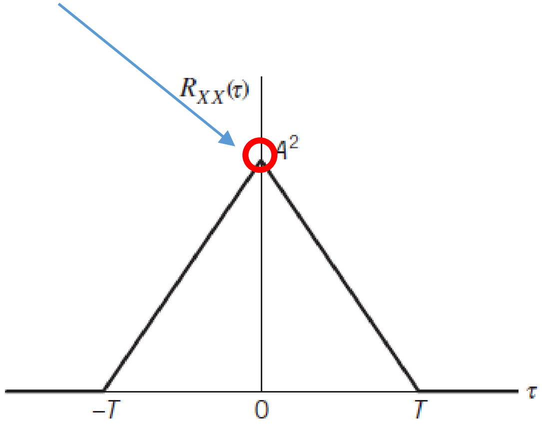
Correlation between two samples of random process X(t)

Autocorrelation function

$$R_{XX}(\tau) = \mathbb{E}[X(t+\tau)X(t)]$$

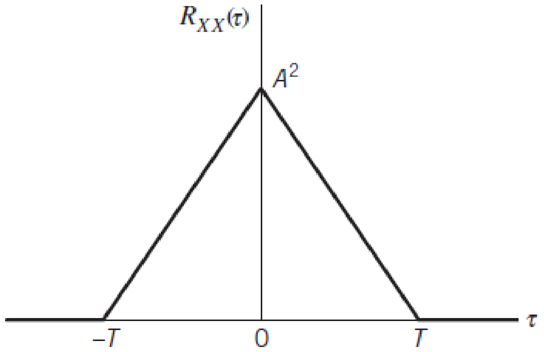
Important properties of the autocorrelation function

$$R_{XX}(0) = \mathbb{E}[X^2(t)]$$



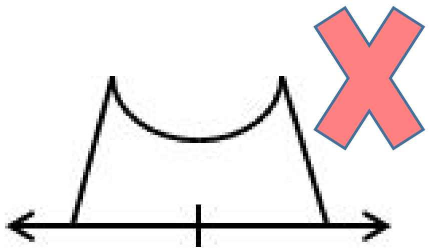
It can be used to calculate power of the signal

$$R_{XX}(\tau) = R_{XX}(-\tau)$$



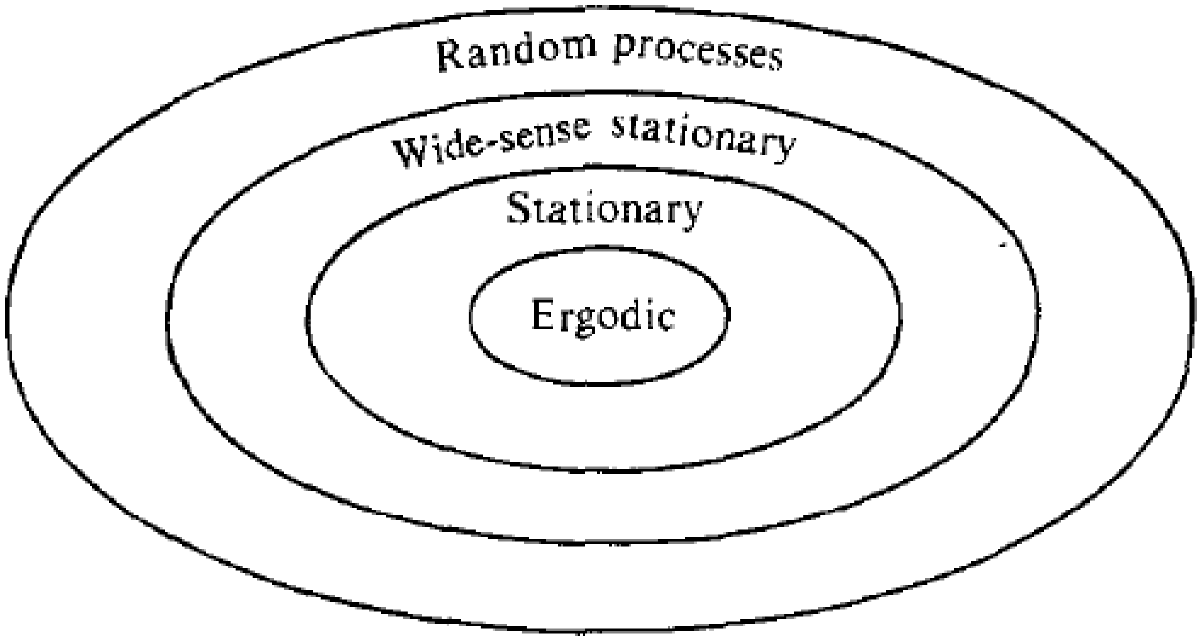
It is symmetric

$$\left|R_{XX}(\tau)\right| \leq R_{XX}(0)$$

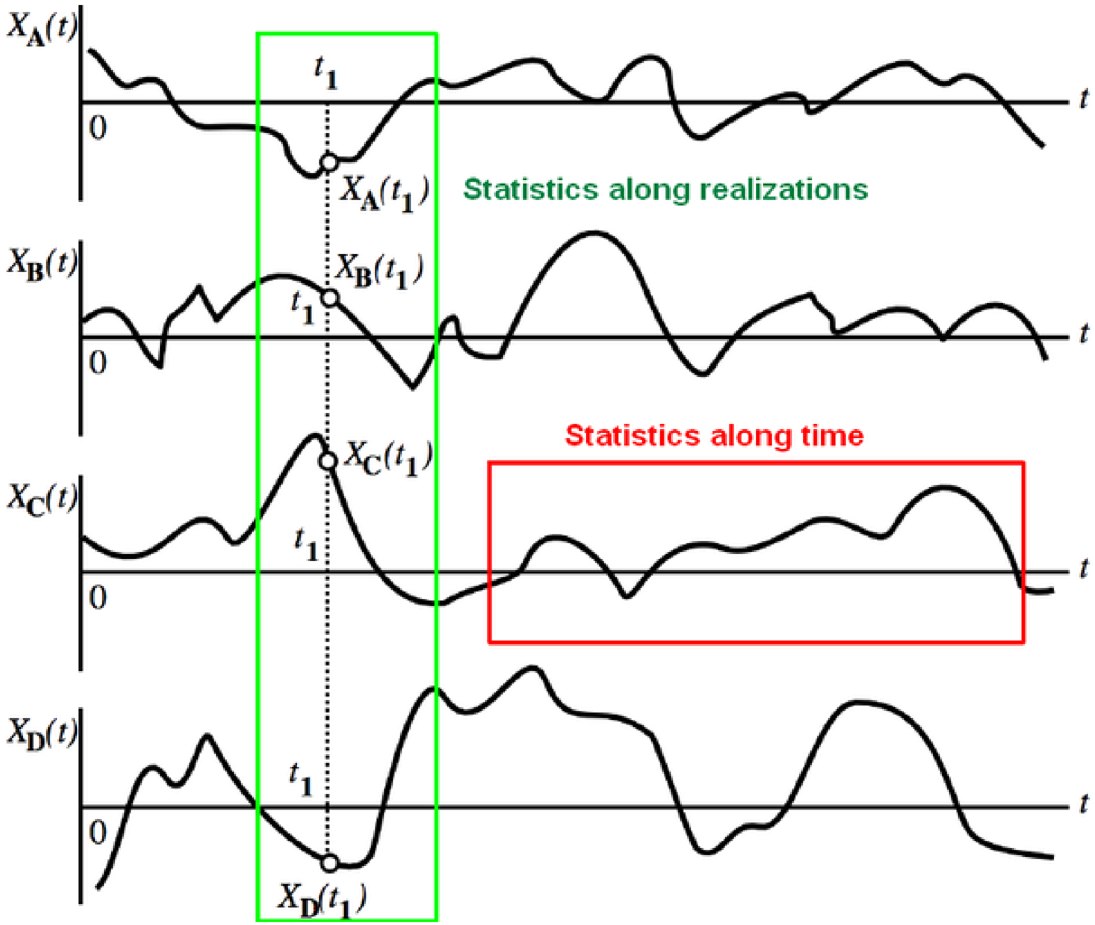


At no point it exceeds the average power

Ergodic Processes



Ergodic Processes



A stationary random process is said to be “ergodic in the mean” if

$$\textbf{\textit{ensamble mean}} = \textbf{\textit{time mean}}$$

Questions?? Thoughts??

EE 302
Probabilistic Methods in Electrical Engineering

with

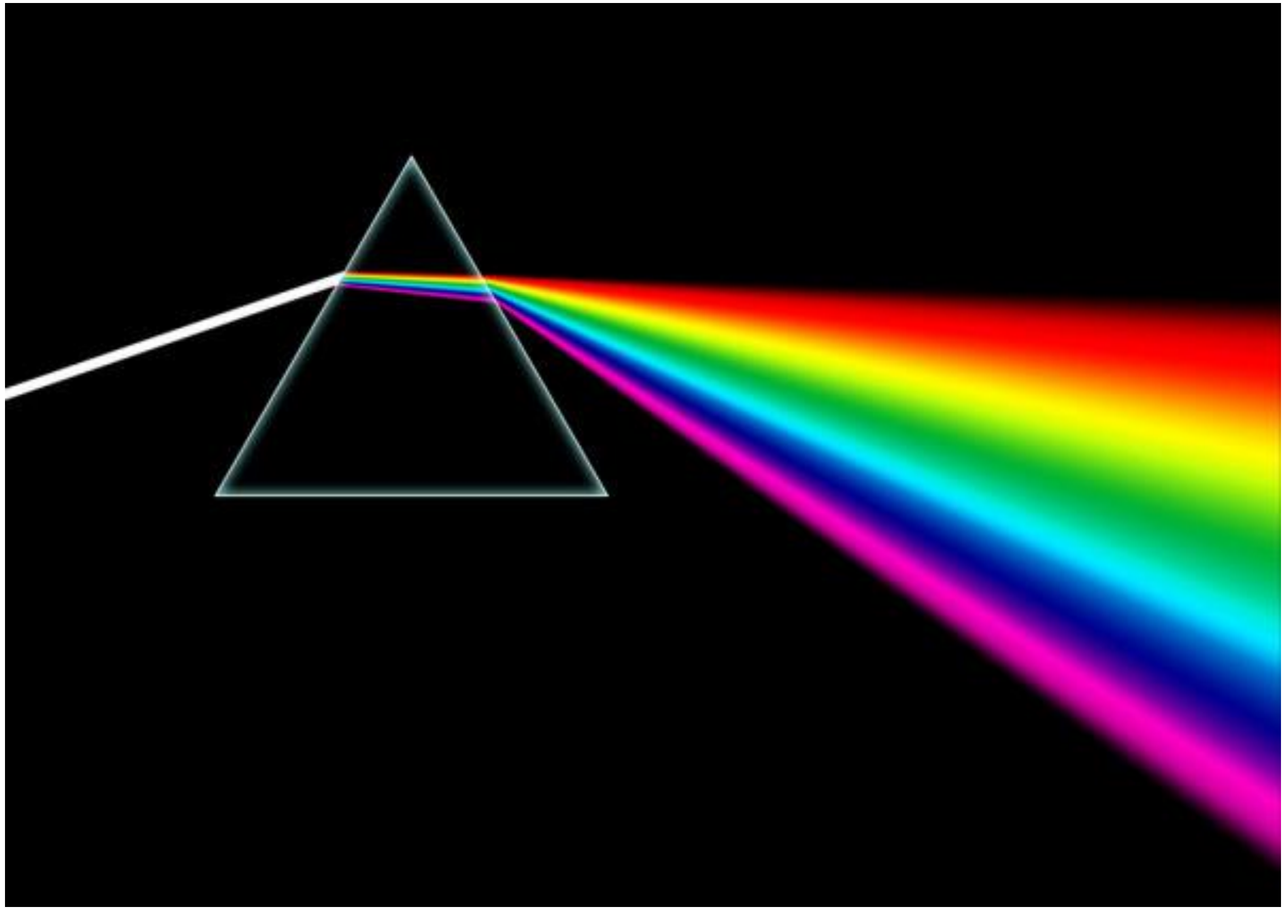
Dr. Naveed R. Butt

@

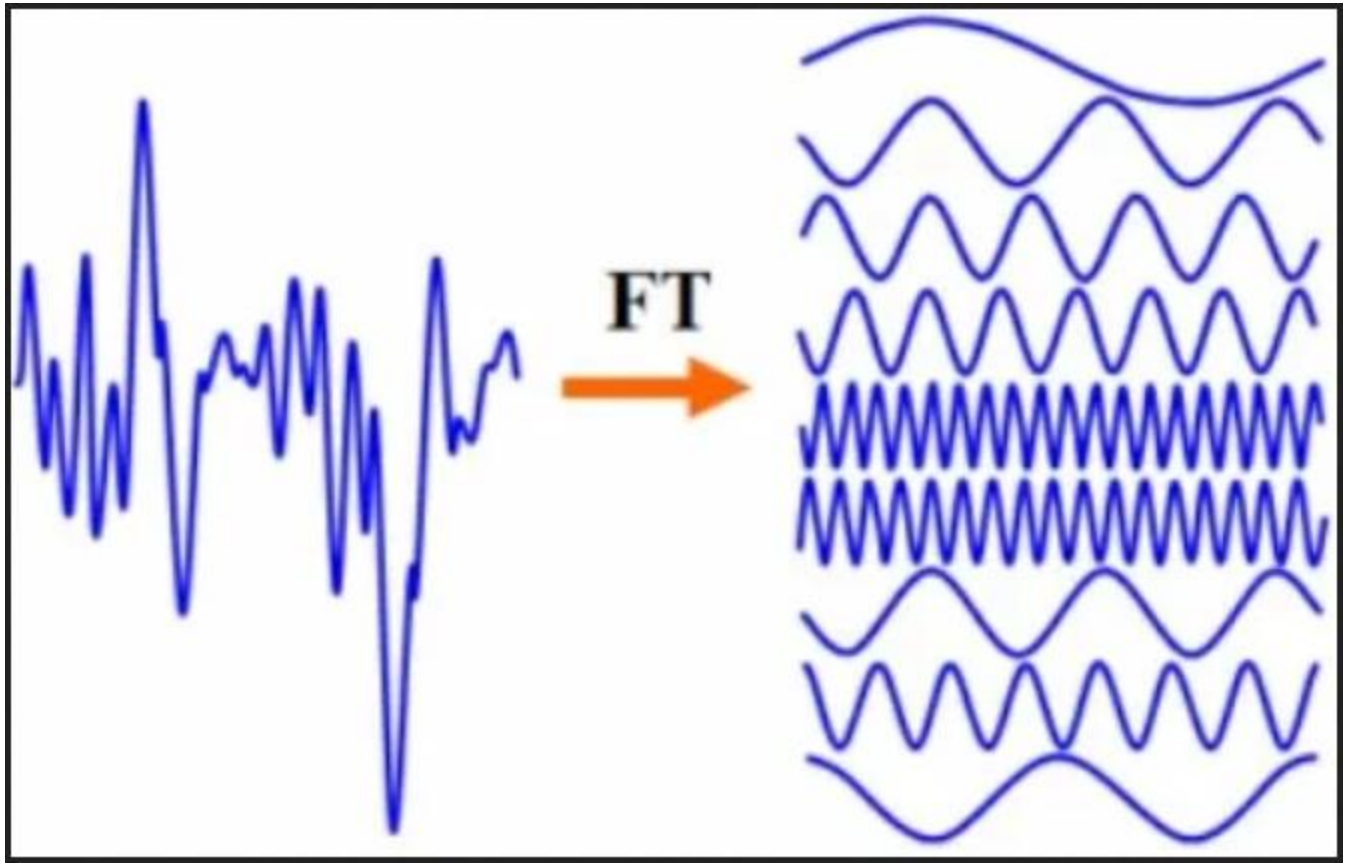
Jouf University

Fourier Transform

Spectra – *the Ghosts in Your Signal*

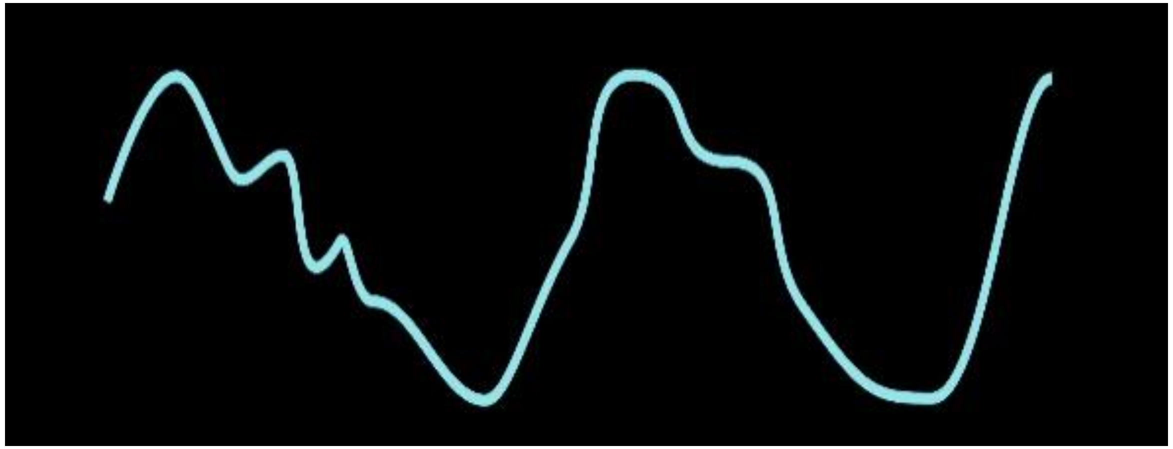


Q. Can we write signals as sums of periodic functions (frequencies)?

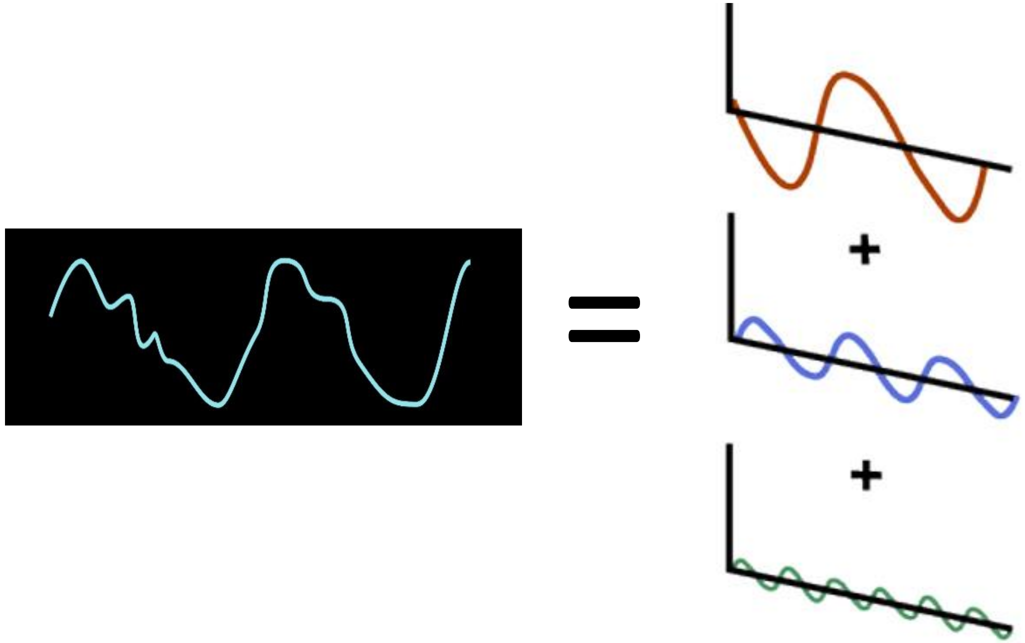


Baking a Fourier Cake

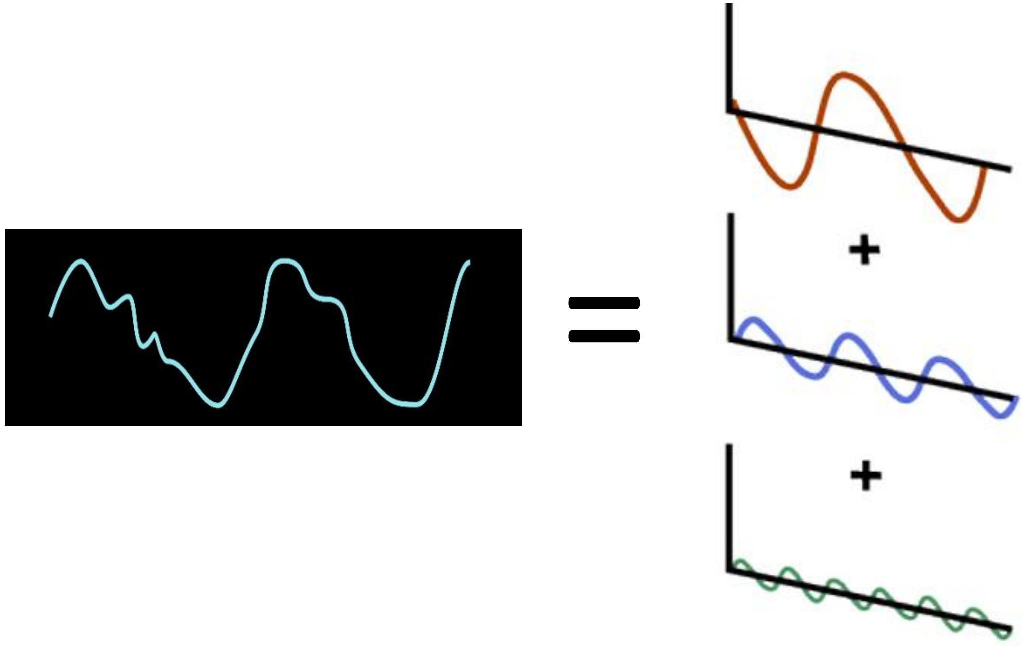
- **Given**: Signal shape (time-domain)
- **Ingredients**: Sinusoids of different frequencies
- **Choose**: How much of the each ingredient (sinusoid) to use?



- In Fourier Transform, we want to look at signals in terms of a fixed set of ingredients
  - Ingredients : Sinusoids of different frequencies

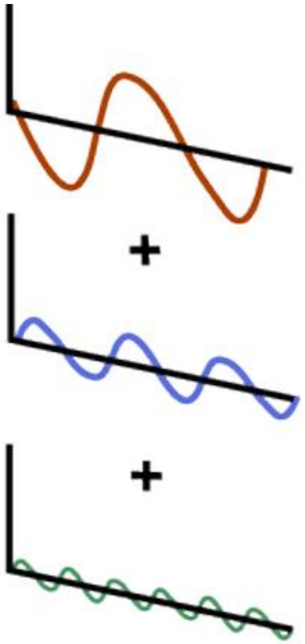


- In Fourier Transform, we want to look at signals in terms of a fixed set of ingredients
  - Ingredients : Sinusoids of different frequencies

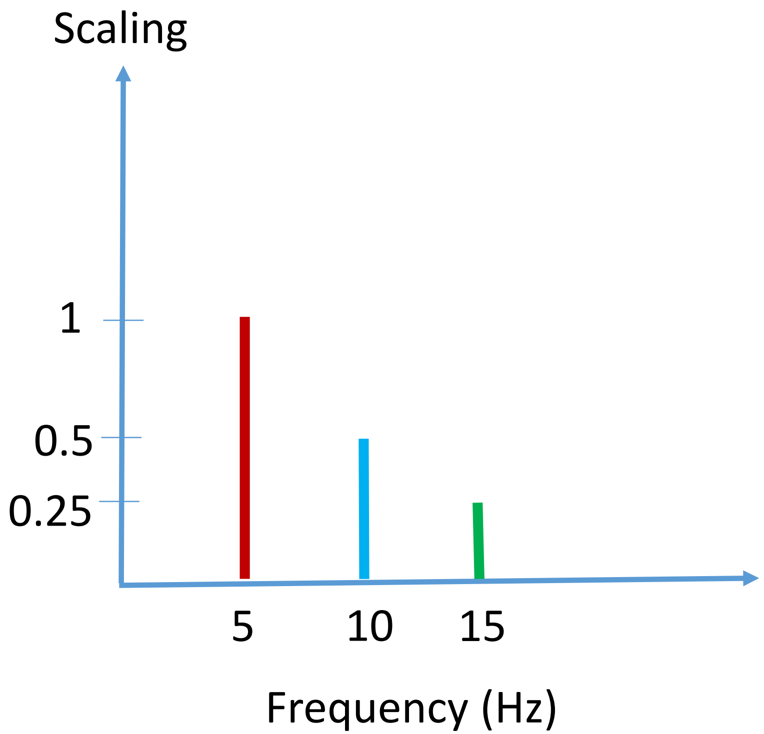


Ingredient (sinusoid frequency)	Amount (scaling)	Process
f_1	1	**Add all**
f_2	0.5	
f_3	0.25	

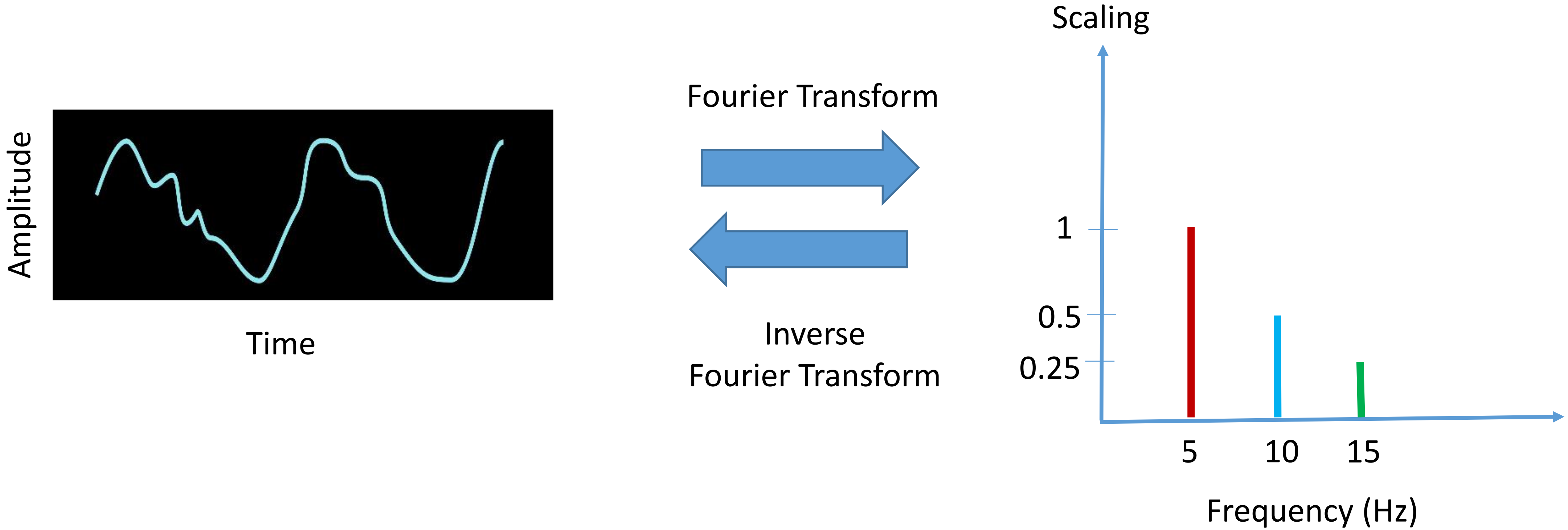
- How is this shown after Fourier transform?



Ingredient (sinusoid frequency)	Amount (scaling)	Process
5 Hz	1	**Add all**
10 Hz	0.5	
15 Hz	0.25	

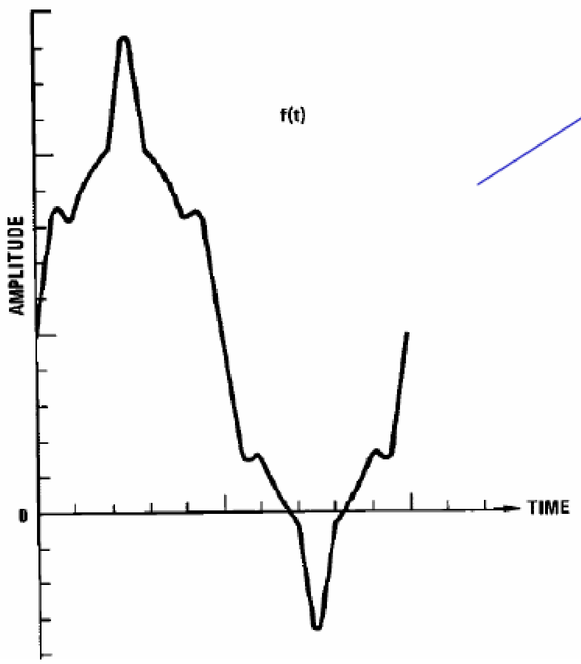


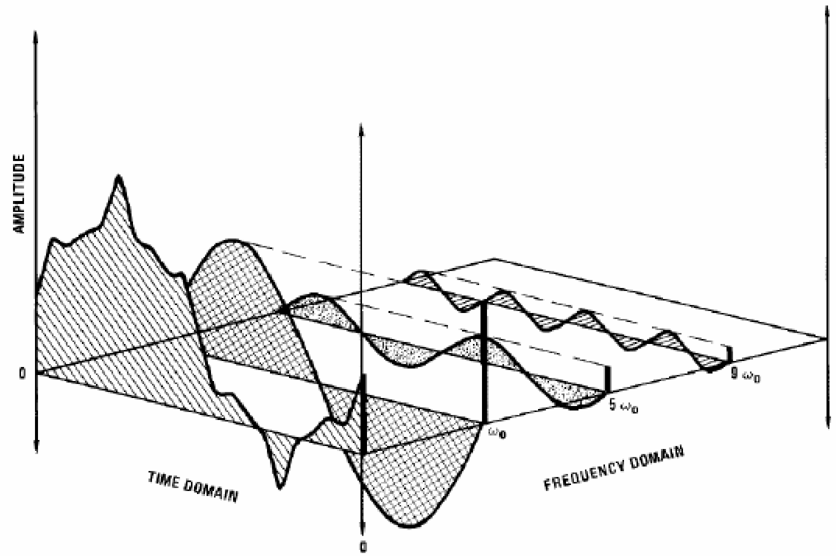
- We mostly skip the middle steps

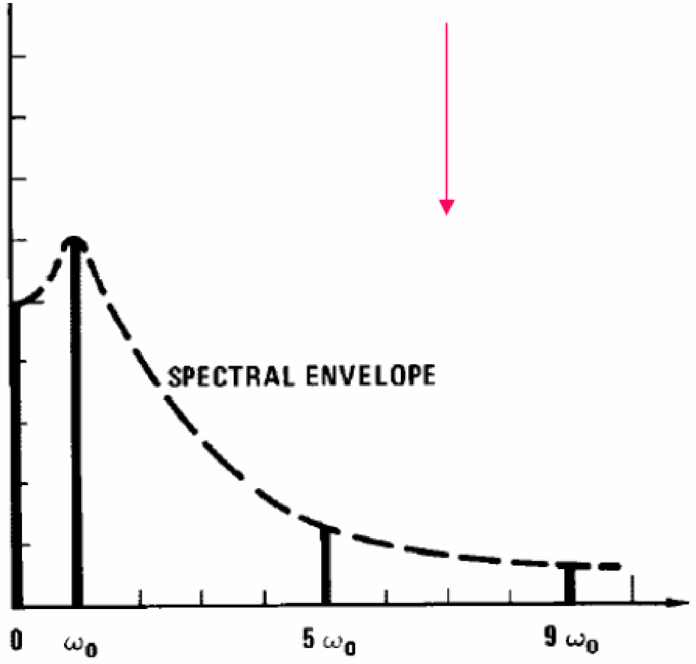


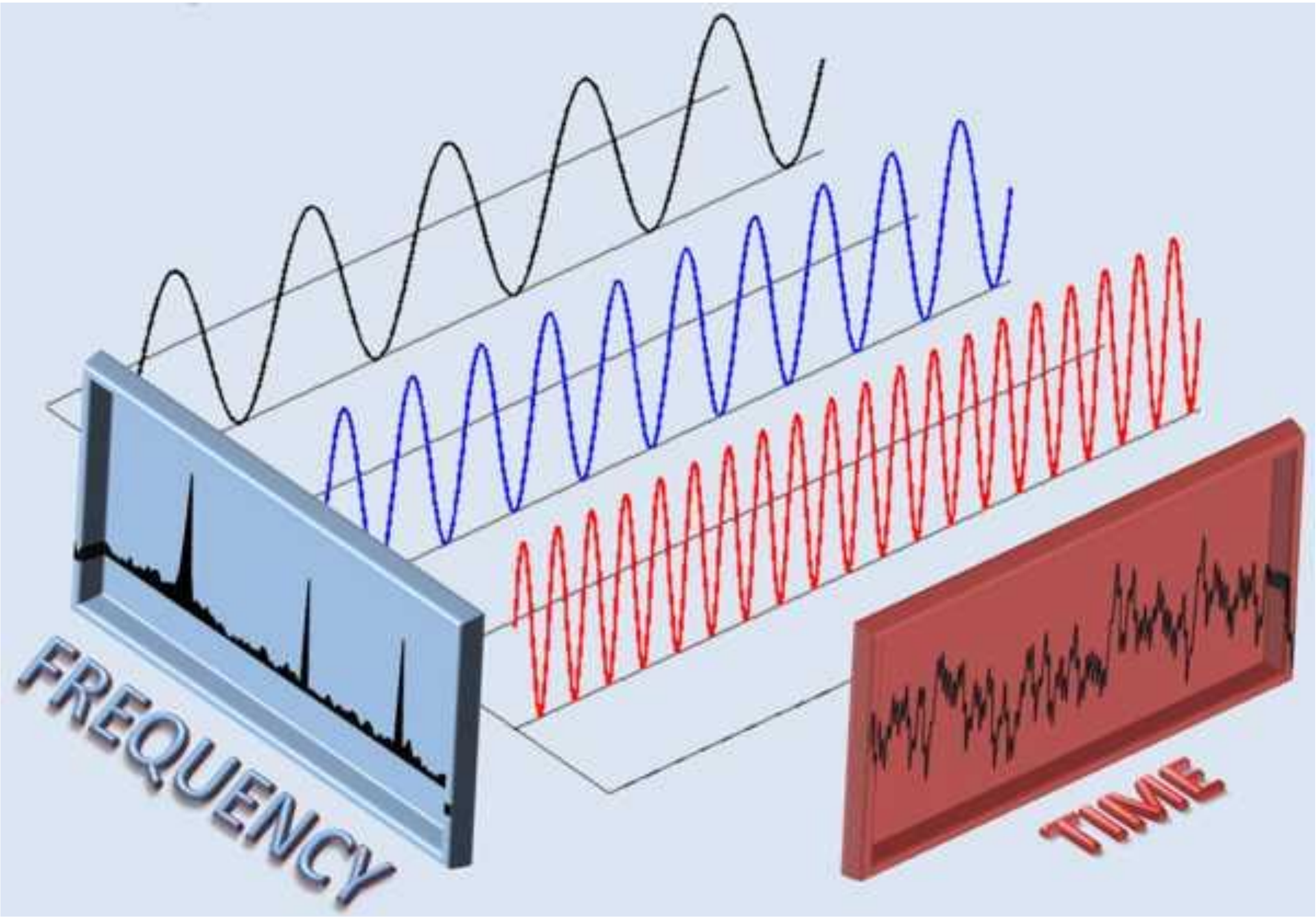
Fourier Transform

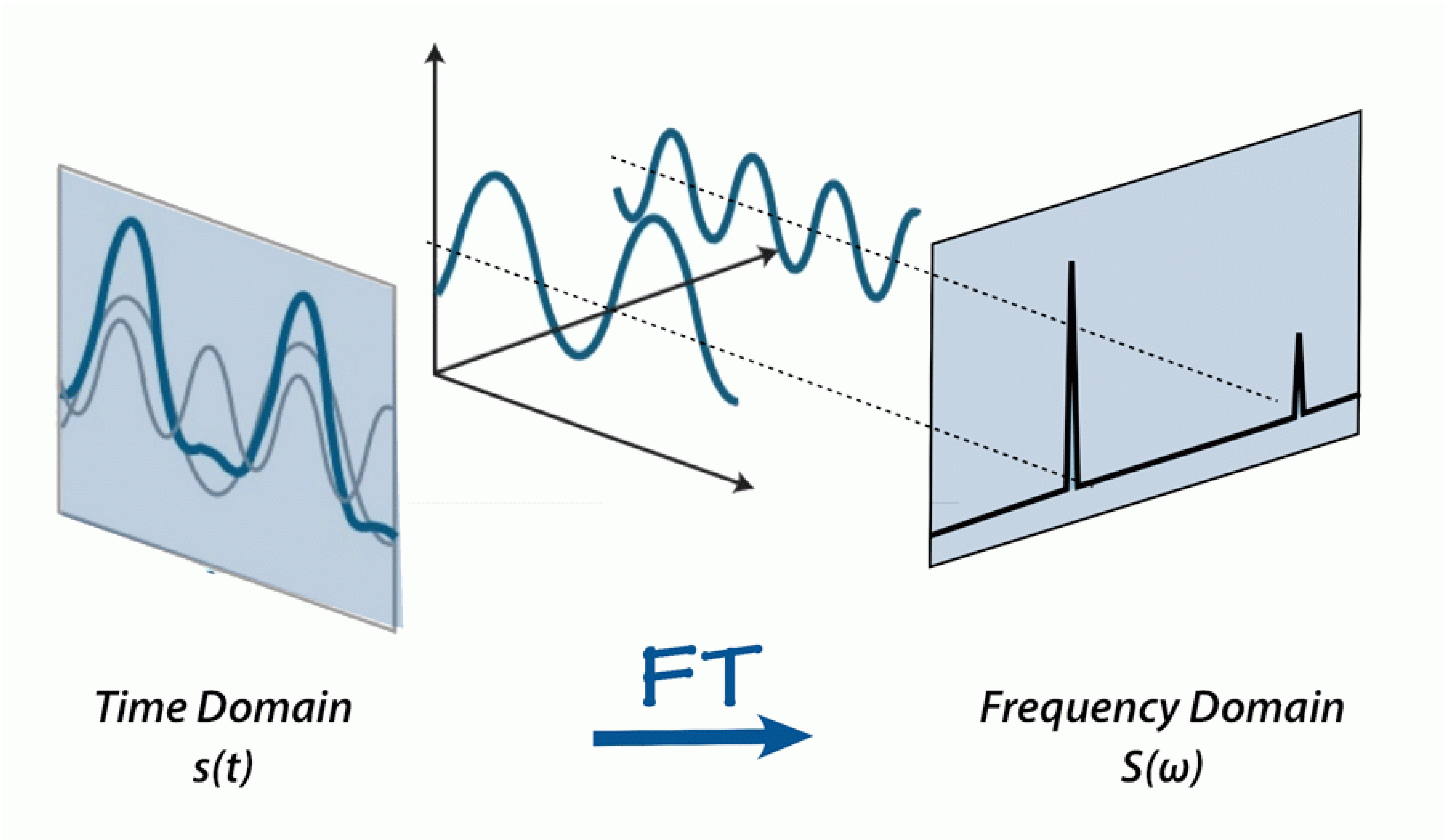
Inverse
Fourier Transform

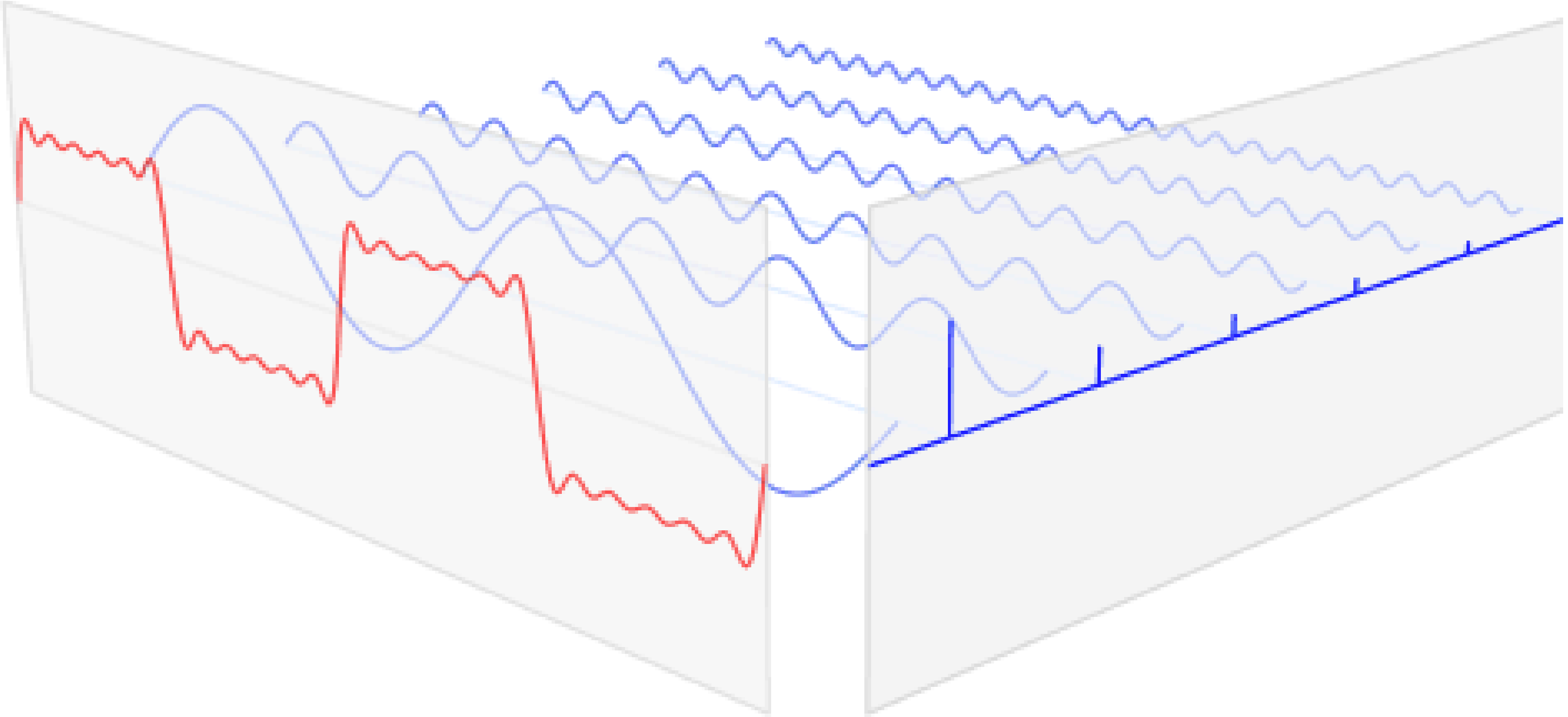

f(t)
AMPLITUDE
0
TIME

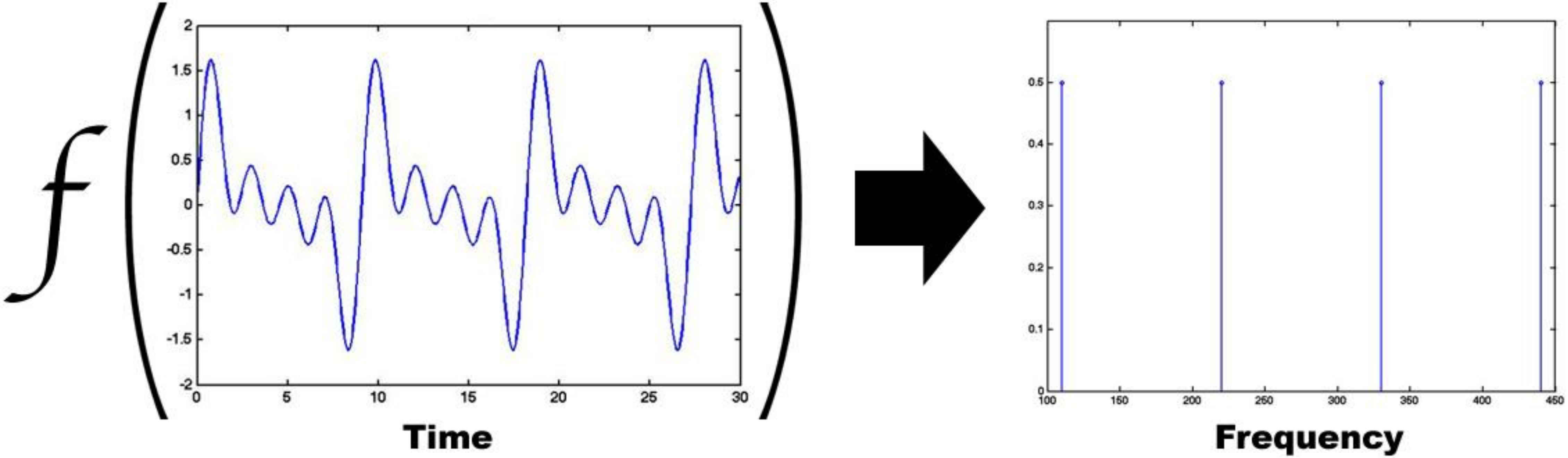

AMPLITUDE
0
9 ω_0
5 ω_0
ω_0
TIME DOMAIN
FREQUENCY DOMAIN
0


SPECTRAL ENVELOPE
0
ω_0
5 ω_0
9 ω_0


FREQUENCY
TIME


Time Domain
s(t)
FT
Frequency Domain
S(ω)

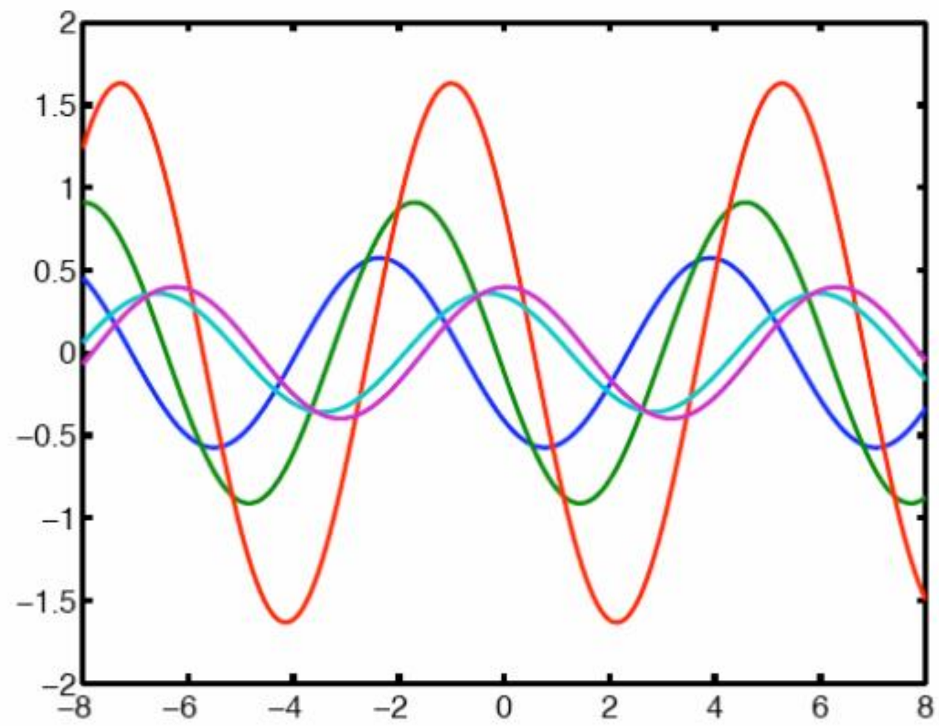



f
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
5
10
15
20
25
30
Time
0.5
0.4
0.3
0.2
0.1
0
100
150
200
250
300
350
400
450
Frequency

Power Spectrum

Frequency content of a random process

- We’ve seen that
  - Fourier Transform can be used to see the “frequency content” of a signal
- But what if the signal is random?
  - **Problem**: frequency content may change from one realization to another!



$$X(t) = A\cos(\omega t + \phi).$$

A and ϕ random variables

Autocorrelation Function to the rescue!!

- We've seen that
  - Fourier Transform can be used to see the "frequency content" of a signal
- But what if the signal is random?
  - **Problem**: frequency content may change from one realization to another!
- Two important observations can help us out
  - 1. The autocorrelation function of a weakly stationary process remains "fixed" between realizations
  - 2. The autocorrelation function contains the same frequencies as the original signal (with "average power" scalings)
- Solution:
  - For a random process it is better to **take the Fourier Transform of the covariance function**

For a random process it is better to take the Fourier Transform of the autocorrelation function!!

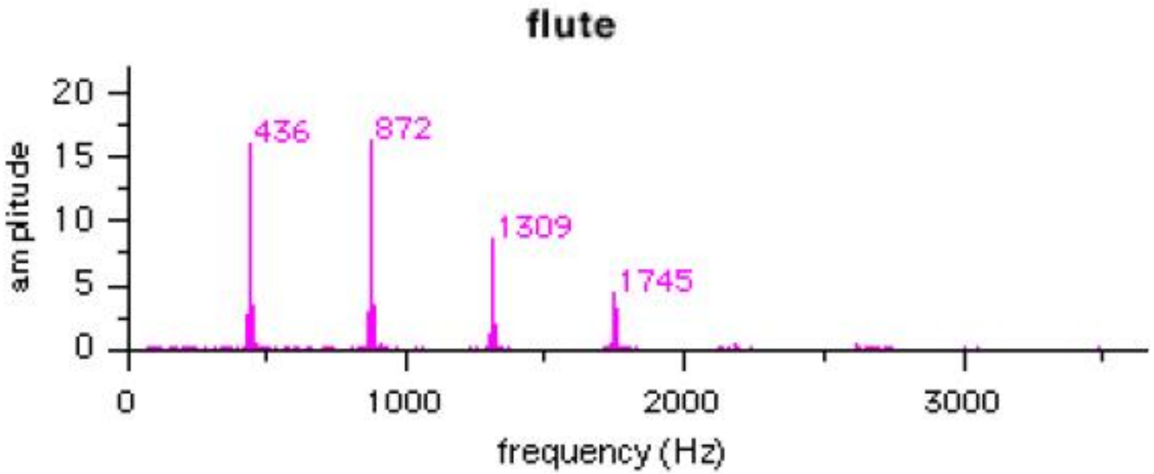
- **Power Spectrum** = Fourier Transform of the autocorrelation function
- Mathematically speaking …

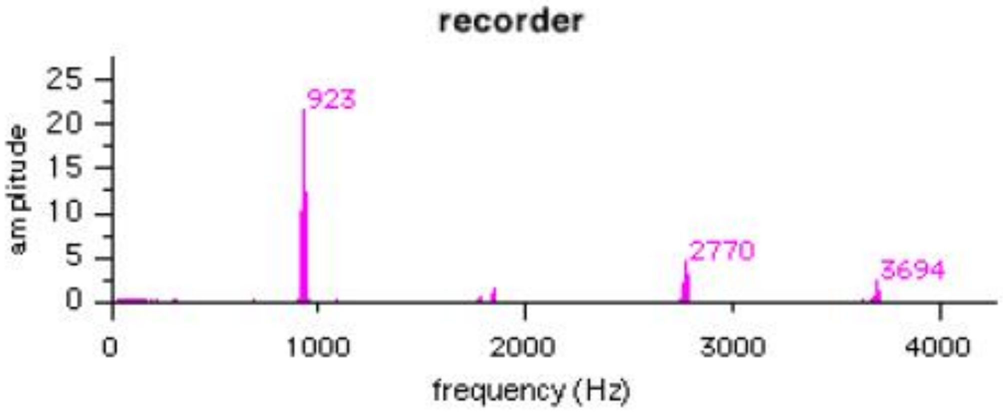
$$S_{XX}(f) = \int_{-\infty}^{\infty} R_{XX}(\tau) \exp(-j2\pi f \tau)\, d\tau$$

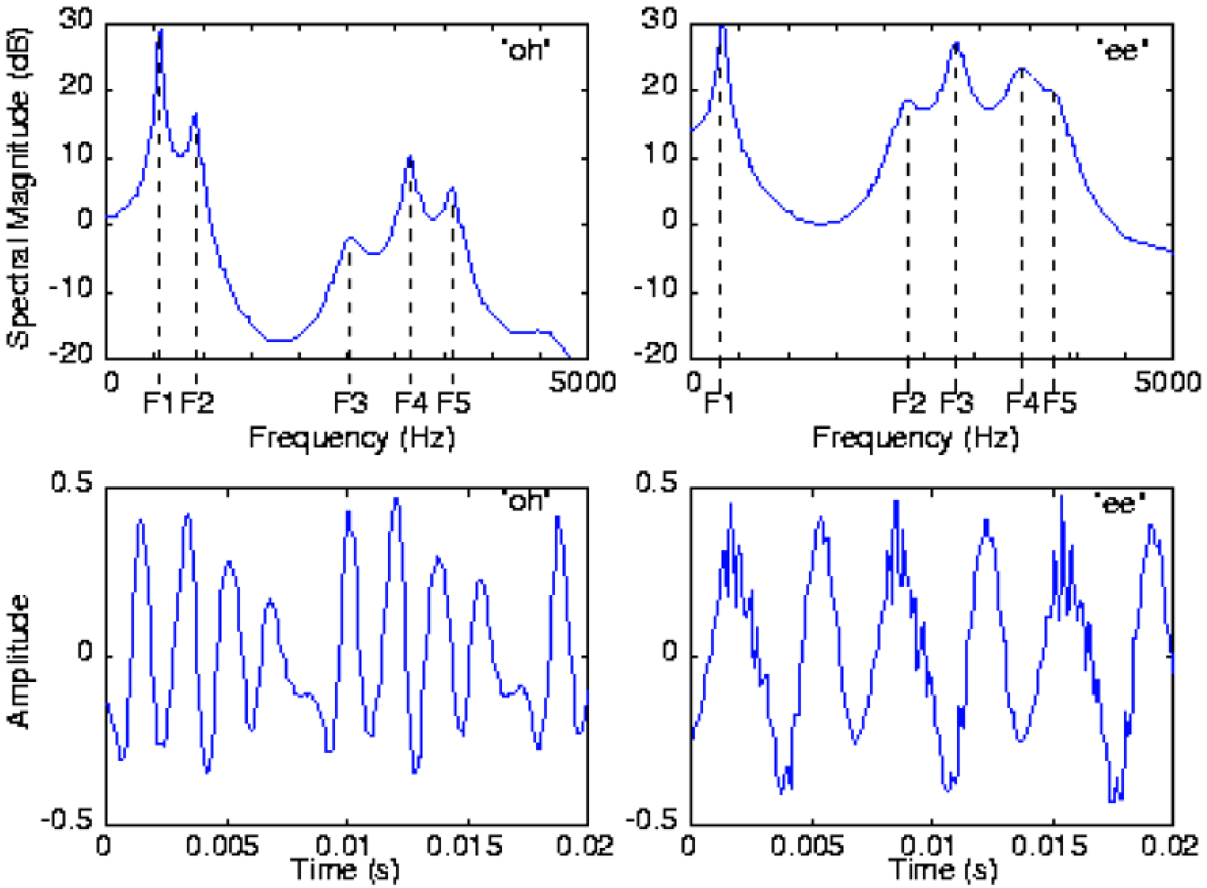
$$R_{XX}(\tau) = \int_{-\infty}^{\infty} S_{XX}(f) \exp(j2\pi f \tau)\, df$$

Recall…

$$R_{XX}(\tau) = \mathbb{E}[X(t+\tau)X(t)]$$


flute
amplitude
20
15
10
5
0
436
872
1309
1745
0
1000
2000
3000
frequency (Hz)


recorder
amplitude
25
20
15
10
5
0
923
2770
3694
0
1000
2000
3000
4000
frequency (Hz)



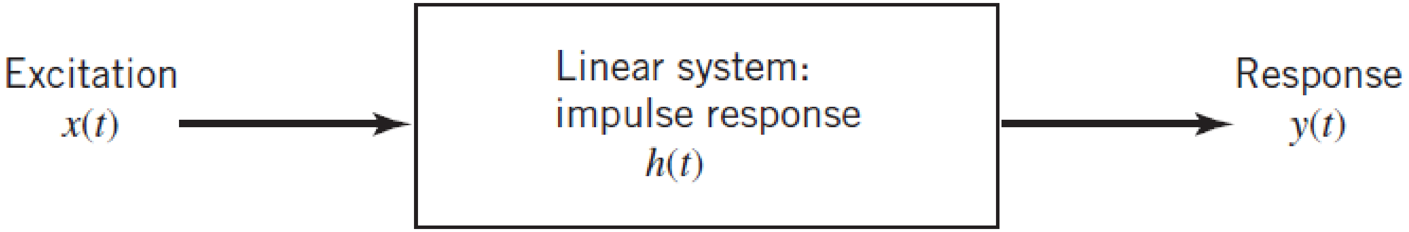
Voice Recognition

Linear Systems

What happens to a signal as it passes through a system?

System = ***channel, filter, etc.***

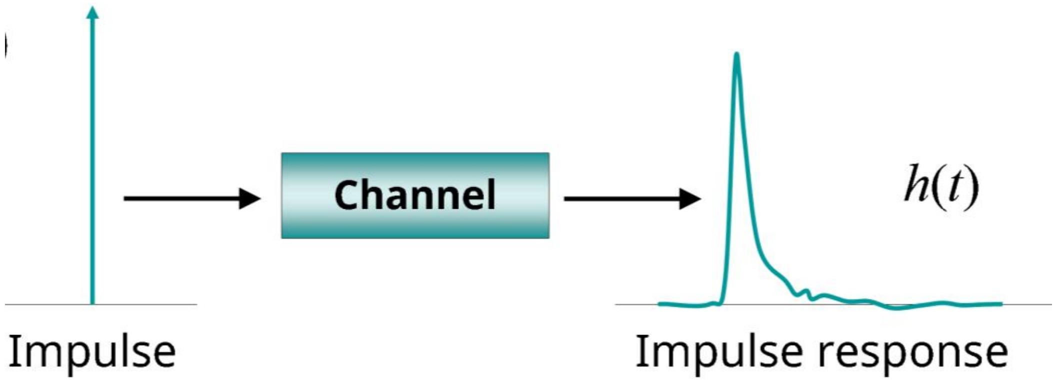
- We shall only consider a special type of systems called **LTI** (linear time-invariant systems).
- We shall only consider weakly stationary signals.



What happens to a signal as it passes through a system?

System = ***channel, filter, etc.***

- An LTI system is fully characterized by its **Impulse Response** (IR)
  - IR = what output the system gives when the input is an impulse (a theoretical sharp pulse)

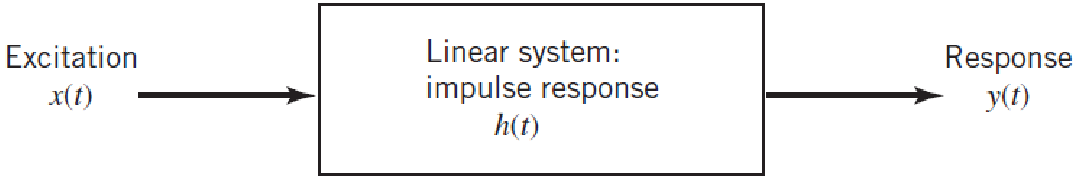


What happens to a signal as it passes through a system?

System = ***channel, filter, etc.***

- If we know the **impulse response** of an LTI system we can find its output to any signal by using **convolution**

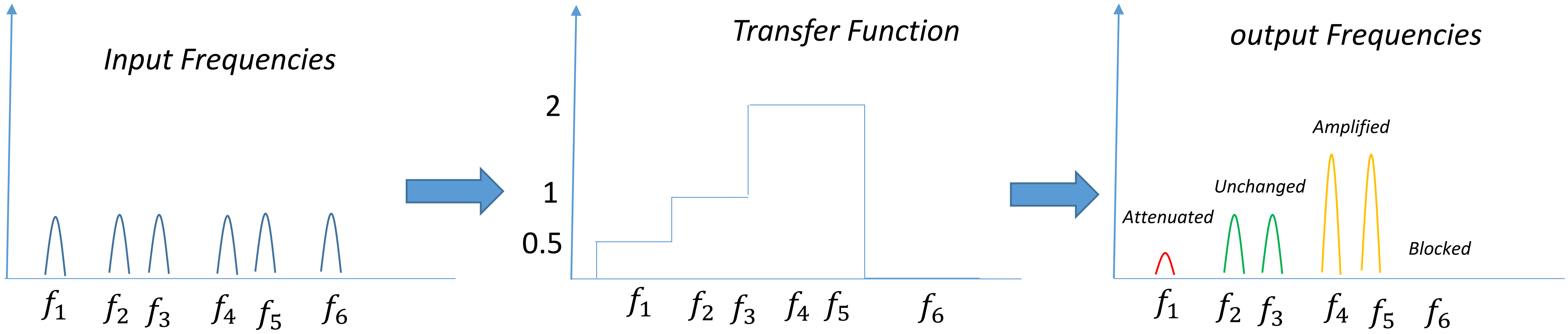
$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,d\tau$$



What happens to a signal as it passes through a system?

System = ***channel, filter, etc.***

- LTI system can also be fully characterized by its **Transfer Function** (TF)
  - TF = *Fourier Transform of Impulse Response*
  - TF = what the system does to different frequencies of the signal (e.g., blocks, allows unchanged, amplifies, attenuates)

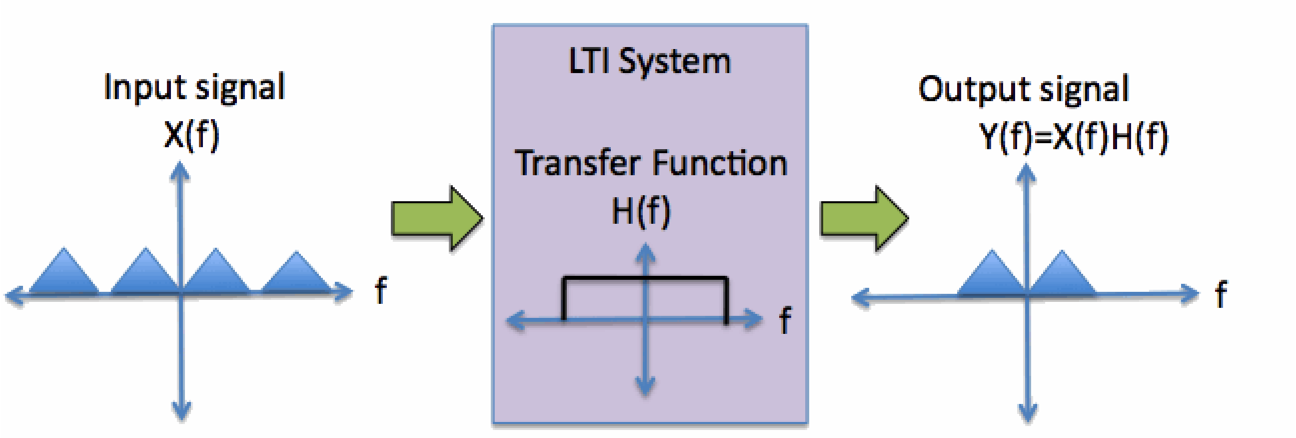


What happens to a signal as it passes through a system?

System = ***channel, filter, etc.***

- If we know the **Transfer Function** of an LTI system we can find what it does to the frequencies of a **deterministic signal** by using simple multiplication in frequency domain

$$Y(f) = H(f)X(f)$$



What happens to a signal as it passes through a system?

System = ***channel, filter, etc.***

- If we know the **Transfer Function** of an LTI system we can find what it does to the **power spectrum** of weakly stationary **random signal** by

$$S_{YY}(f) = |H(f)|^2 S_{XX}(f)$$

$$|H(f)|^2 = H(f)H^*(f)$$

Questions?? Thoughts??